Springer Series in Optical Sciences 206

Mario Bertolotti
Concita Sibilia
Angela M. Guzmán

Evanescent Waves in Optics

An Introduction to Plasmonics

Springer

Springer Series in Optical Sciences

Volume 206

Founded by

H. K. V. Lotsch

Editor-in-Chief

William T. Rhodes, Georgia Institute of Technology, Atlanta, USA
Florida Atlantic University, Boca Raton, USA

Editorial Board

Ali Adibi, Georgia Institute of Technology, Atlanta, USA
Theodor W. Hänsch, Max-Planck-Institut für Quantenoptik, Garching, Germany
Ferenc Krausz, Lehrstuhl für Experimentelle Physik, Garching, Germany
Max-Planck-Institut für Quantenoptik, Garching, Germany
Barry R. Masters, Cambridge, USA
Katsumi Midorikawa, Saitama, Japan
Herbert Venghaus, Heinrich-Hertz-Institut, Berlin, Germany
Horst Weber, Technische Universität Berlin, Berlin, Germany
Harald Weinfurter, Ludwig-Maximilians-Universität München, München, Germany

Springer Series in Optical Sciences

The Springer Series in Optical Sciences, under the leadership of Editor-in-Chief William T. Rhodes, Georgia Institute of Technology and Florida Atlantic University, USA, provides an expanding selection of research monographs in all major areas of optics: lasers and quantum optics, ultrafast phenomena, optical spectroscopy techniques, optoelectronics, quantum information, information optics, applied laser technology, industrial applications, and other topics of contemporary interest.

With this broad coverage of topics, the series is of use to all research scientists and engineers who need up-to-date reference books.

The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the Editor-in-Chief or one of the Editors. See also www.springer.com/series/624

Editor-in-Chief

William T. Rhodes
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250
USA

Florida Atlantic University
Boca Raton, USA
e-mail: bill.rhodes@ece.gatech.edu

Editorial Board

Ali Adibi
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250
USA
e-mail: adibi@ee.gatech.edu

Theodor W. Hänsch
Max-Planck-Institut für Quantenoptik
Hans-Kopfermann-Straße 1
85748 Garching, Germany
e-mail: t.w.haensch@physik.uni-muenchen.de

Ferenc Krausz
Ludwig-Maximilians-Universität München
Lehrstuhl für Experimentelle Physik
Am Coulombwall 1
85748 Garching, Germany

Max-Planck-Institut für Quantenoptik
Hans-Kopfermann-Straße 1
85748 Garching, Germany
e-mail: ferenc.krausz@mpq.mpg.de

Barry R. Masters
Cambridge, USA

Katsumi Midorikawa
Saitama, Japan

Herbert Venghaus
Fraunhofer Institut für Nachrichtentechnik
Heinrich-Hertz-Institut
Einsteinufer 37
10587 Berlin, Germany
e-mail: venghaus@hhi.de

Horst Weber
Optisches Institut
Technische Universität Berlin
Straße des 17. Juni 135
10623 Berlin, Germany
e-mail: weber@physik.tu-berlin.de

Harald Weinfurter
Sektion Physik
Ludwig-Maximilians-Universität München
Schellingstraße 4/III
80799 München, Germany
e-mail: harald.weinfurter@physik.uni-muenchen.de

More information about this series at http://www.springer.com/series/624

Mario Bertolotti · Concita Sibilia
Angela M. Guzmán

Evanescent Waves in Optics

An Introduction to Plasmonics

Springer

Mario Bertolotti
"La Sapienza" SBAI
University of Rome
Rome
Italy

Concita Sibilia
"La Sapienza" SBAI
University of Rome
Rome
Italy

Angela M. Guzmán
Physics Department
National University of Colombia
Bogotá
Colombia

ISSN 0342-4111 ISSN 1556-1534 (electronic)
Springer Series in Optical Sciences
ISBN 978-3-319-61260-7 ISBN 978-3-319-61261-4 (eBook)
DOI 10.1007/978-3-319-61261-4

Library of Congress Control Number: 2017954456

© Springer International Publishing AG 2017
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The quest for ever smaller optical systems has brought us to devices of the dimension of a wavelength or less, down to nanometres. Far-field optics is to a great extent incompatible with structures of that size, and one must thus turn to near-field optics and surface evanescent waves, optical waves that propagate only along, and whose energy is spatially concentrated in the immediate vicinity of a surface.

The basic nature of evanescent waves has been known for decades, but they have found little application until recently. We now know that these waves allow the operation of a large number of devices, and that they can be employed in overcoming the classical Rutherford-Abbe's resolution limit.

Evanescent waves travelling on the surface of metals or semiconductors are called plasmons. Plasmons seem to be the best solution for the building and management of nanoscale devices in which, for example, light is coupled into and out of an optical fibre and shrink in a structure much smaller than the vacuum wavelength. The new discipline of *plasmonics* encompasses the generation and use of plasmons both as a travelling excitation in a nanostructure and as a stationary enhancement of the electrical field near metal nanosurfaces.

In this short monograph we present an introduction to evanescent waves and plasmons, discussing their properties and uses, and showing how they are of fundamental importance to the operation and understanding of nanoscale optical devices.

Unlike some of the specialized textbooks listed in the references, this book is not exhaustive in its treatment of plasmonics; rather, it is intended to provide the reader with a general survey of the field and the state of the art, including the underlying physics and applications of evanescent waves and plasmons. It is a first approach, not an exhaustive textbook.

The book is organized so as to offer in Chap. 1, a brief review of the basic concepts of electromagnetism that will be needed in subsequent chapters. The second chapter introduces evanescent waves in the most natural way through reflection and refraction and shows how they appear in diffraction problems. In the third chapter, the important role that evanescent waves play in optical waveguides and sensors is discussed. The application of evanescent waves in super-resolution devices is then briefly presented in chap. IV. Plasmons are the most important manifestation of

evanescent waves and form the subject of chap. V, while some of their applications are presented in chap. VI. In working with very small devices, one must always bear in mind that quantum effects may play a significant role in determining device behaviour. The last chapter, therefore, discusses quantization of evanescent waves and the possibility of single photon plasmonic devices.

The book is intended for students and researchers who wish to enter the field of plasmonics or to gain insight into the subject and can serve as an introduction to more complete and in-depth discussions that can be found in many excellent textbooks.

Two of the authors (MB and CS) wish to thank the help of Alessandro Belardini for reading the manuscript and making some figures.

The work and help of the whole staff of Springer is gratefully acknowledged.

Rome, Italy, May 8, 2017

Mario Bertolotti
Concita Sibilia
Angela M. Guzmán

Contents

1 Basic Electromagnetism 1
1.1 Introduction 1
1.2 Maxwell Equations 1
1.3 Waves 2
1.4 Phase Velocity 4
1.5 Dispersion 5
1.6 Drude-Lorentz Model for Metals 9
1.7 Pulses and Group Velocity 14
1.8 Polarization 16
1.9 Jones Matrices, Stokes Parameters and the Poincarè Sphere 19
1.10 Optically Anisotropic Media 23
1.11 Chirality 29
1.12 Gaussian Beams 31

2 Evanescent Waves 35
2.1 Introduction 35
2.2 Reflection and Refraction 35
2.3 Evanescent Waves 45
2.4 Energy Transport by Evanescent Waves 49
2.5 Tunnelling Effect 50
2.6 Reflection and Refraction in the Presence of Absorption 51
2.7 Reflection and Refraction with Materials with Negative Refractive Index 55
2.8 X-ray Evanescent Waves 57
2.9 Reflection and Refraction of Plane Waves at a Boundary Between an Isotropic and a Birefringent Medium 59
2.10 The Plane Wave Decomposition of a Field 60
2.11 The Classical Limit of Resolution Explained 62
2.12 Reflection and Refraction of Gaussian Beams 64
2.13 Evanescent Waves in Diffraction 65

3 Evanescent Waves in Optical Waveguides 69
3.1 Introduction 69
3.2 Planar Waveguides 69
3.3 Coupling of Light to a Planar Waveguide 81
3.4 Coupling of Two Waveguides. 86
3.5 Optical Fibres. 88
3.6 Multilayers and PBG 93
3.7 The Role of Evanescent Waves in Waveguide Sensors 106

4 High Resolution Optical Microscopes 111
4.1 Introduction 111
4.2 Scanning Near-field Optical Microscopy (SNOM). 112
4.3 Scanning Tunnelling Optical Microscope (STOM) 118
4.4 Total Internal Reflection Fluorescence (TIRF). 122

5 Plasmons 127
5.1 Introduction 127
5.2 Plasmon Solutions 128
5.3 Bulk Plasmons 130
5.4 Surface Plasmon Polaritons (SPPs) 131
5.5 Properties of Plasmons 136
5.6 Excitation and Coupling of Plasmons 139
5.7 Multilayer Systems. 144
5.8 Localized Surface Plasmons 150
5.9 Surface Phonon Polaritons in Dielectrics and Semiconductors . . . 156
5.10 The Plasmons in Optical Nonlinear Materials 160
5.11 Other Surface Waves 162

6 Applications of Plasmons 169
6.1 Introduction 169
6.2 Surface Enhanced Raman Scattering (SERS). 170
6.3 Surface Plasmon Sensors 173
6.4 Extraordinary Optical Transmission Through Arrays of Sub-wavelength Holes. 177
6.5 Surface Plasmon Circuitry 182
6.6 Plasmon Lasers and SPASER 188
6.7 Plasmons for Solar Cells. 198
6.8 Plasmon Microscopy 201
6.9 Black-body Spatial and Temporal Coherence 202
6.10 Controlled Thermal Emission Using Plasma Resonances. 208

7 Quantization of Evanescent Waves. 209
7.1 Introduction 209
7.2 Quantization of the Electromagnetic Field in One Dimension . . . 211
7.3 Quantum States of the Electromagnetic Field 215
7.4 Quantization of the Electromagnetic Field in Lossless Dielectric Media. 219

7.5 What are photons? 226
7.6 The Problem of Localizing Photons 228
7.7 Expansion of the Field and Their Orthonormalization 229
7.8 Quantization of Evanescent Waves 232
7.9 Plasmons in Bulk Metals 239
7.10 Surface Plasmon Polaritons 241
7.11 Localized Surface Plasmon Resonances 248
7.12 Absorption of Evanescent Photons and Stimulated Emission of Surface Plasmons 250
7.13 Plasmons and Quantum Information 251

Index 257

Chapter 1
Basic Electromagnetism

1.1 Introduction

We assume the reader is already familiar with the main subjects of electromagnetism and therefore limit ourselves in this chapter to a brief resumè of some basic concepts that will be used later in the book. The most important subject here discussed is paragraph 1.5 on dispersion in which we introduce the classical Drude formula and the modifications needed to treat with metals. Dispersion in metals is basic for plasmons.

1.2 Maxwell Equations

We consider Maxwell's equations of macroscopic electromagnetism in the following form

$$\nabla \cdot \mathbf{D} = \rho_{\text{ext}} \tag{1.1a}$$

$$\nabla \cdot \mathbf{B} = 0 \tag{1.1b}$$

$$\nabla \text{x} \mathbf{E} = -\partial \mathbf{B} / \partial t \tag{1.1c}$$

$$\nabla \text{x} \mathbf{H} = \mathbf{j}_{\text{ext}} + \partial \mathbf{D} / \partial t. \tag{1.1d}$$

These equations link the four macroscopic fields **D** (dielectric displacement), **E** (the electric field), **H** (the magnetic field), and **B** (the magnetic induction) with the external charge and current densities ρ_{ext} and $\mathbf{j}_{\text{ext}}$, respectively.

The four macroscopic fields are further linked via the polarization **P** and magnetization **M** by

$$\mathbf{D} = \varepsilon_o \mathbf{E} + \mathbf{P} \tag{1.2a}$$

$$\mathbf{B} = \mu_o \mathbf{H} + \mathbf{M} \tag{1.2b}$$

© Springer International Publishing AG 2017

M. Bertolotti et al., *Evanescent Waves in Optics*, Springer Series in Optical Sciences,
DOI 10.1007/978-3-319-61261-4_1

where ε_o and μ_o are the *electric permittivity* and *magnetic permeability* of vacuum, respectively.

In linear, isotropic media

$$\mathbf{P} = \varepsilon_o \chi \mathbf{E} \tag{1.3}$$

and

$$\mathbf{D} = \varepsilon_o \varepsilon_r \mathbf{E} = \varepsilon_o \mathbf{E} + \mathbf{P} \tag{1.4a}$$

$$\mathbf{B} = \mu_o \mu_r \mathbf{H} = \mu_o \mathbf{H} + \mathbf{M}. \tag{1.4b}$$

χ is the *electric susceptibility*, $\varepsilon_r = 1 + \chi$ is called the *relative permittivity* and μ_r the *relative permeability*. If the medium is non-magnetic $\mu_r = 1$. For nonlinear media the polarization (1.3) is not proportional to the electric field and can often be written as a power series of the electric field, which gives rise to a number of nonlinear optical processes.[1]

The last constitutive relation is

$$\mathbf{j} = \sigma \mathbf{E}, \tag{1.5}$$

where σ is the *electric conductivity*. This last equation is valid only for ohmic materials. Metals and semiconductors usually satisfy it, but ultrathin metallic films behave in a non-ohmic fashion as their thickness reaches the nanoscale.

1.3 Waves

For an isotropic linear dielectric medium, assuming j_{ext}, $\rho_{ext} = 0$, we obtain the wave equations

$$\nabla^2 \mathbf{E} - (1/v^2)\partial^2 \mathbf{E}/\partial t^2 = 0 \tag{1.6a}$$

$$\nabla^2 \mathbf{H} - (1/v^2)\partial^2 \mathbf{H}/\partial t^2 = 0 \tag{1.6b}$$

where

$$v = 1/(\varepsilon\mu)^{1/2} = 1/[(\varepsilon_o\mu_o)(\varepsilon_r\mu_r)]^{1/2} = c/(\varepsilon_r\mu_r)^{1/2} = c/n \tag{1.7}$$

being $c = 1/(\varepsilon_o\mu_o)^{1/2}$ the light speed in vacuum and $n = (\varepsilon_r\mu_r)^{1/2}$ is called *refractive index*.

[1] See for ex. Shen YR (1984) The principles of nonlinear optics. Wiley

Considering harmonic fields of the form

$$\mathbf{E} = \mathbf{E}_o e^{i\omega t} \quad \text{and} \quad \mathbf{H} = \mathbf{H}_o e^{i\omega t} \tag{1.8}$$

with $\omega = 2\pi\nu$, and ν the frequency, (1.6) become

$$\nabla^2 \mathbf{E_o} + (\omega^2/v^2)\mathbf{E}_o = 0 \tag{1.9a}$$

$$\nabla^2 \mathbf{H_o} + (\omega^2/v^2)\mathbf{H}_o = 0 \tag{1.9b}$$

that are the so-called *Helmholtz equations*.

We consider one solution in the form of plane waves

$$\mathbf{E} = \mathbf{A} e^{i(\omega t - \mathbf{k}\cdot\mathbf{r})} \tag{1.10a}$$

and

$$\mathbf{H} = \mathbf{A}' e^{i(\omega t - \mathbf{k}\cdot\mathbf{r})} \tag{1.10b}$$

which represents a plane wave moving in the direction of positive r (*forward wave*) and

$$\mathbf{E} = \mathbf{A} e^{i(\omega t + \mathbf{k}\cdot\mathbf{r})} \tag{1.10c}$$

and

$$\mathbf{H} = \mathbf{A}' e^{i(\omega t + \mathbf{k}\cdot\mathbf{r})} \tag{1.10d}$$

which represents a wave moving in the direction of decreasing r (*backward wave*).

The vector $\mathbf{k}$ is called the *wave-vector* or *wave number* or *propagation vector* and is directed in the direction of the motion of the wave. Its modulus is $k = 2\pi/\lambda$ where λ is the wavelength in the medium. Solutions of the form (1.10) are called sinusoidal waves. Because $\mathbf{E}$ is a real quantity, one should consider the real part or the imaginary part of the exponential.

The product $\mathbf{k} \cdot \mathbf{r}$ is, in Cartesian coordinates

$$\mathbf{k}\cdot\mathbf{r} = (\mathbf{u_x}k_x + \mathbf{u_y}k_y + \mathbf{u_z}k_z)(\mathbf{u_x}x + \mathbf{u_y}y + \mathbf{u_z}z) = k_x x + k_y y + k_z z \tag{1.11}$$

where $\mathbf{u_i}$'s are the unitary vectors along the three coordinate axis. By substitution of (1.11) into (1.6) we have in vacuum

$$\omega^2/|k_o^2| = c^2 \quad \text{and} \quad k_o = \pm\omega/c = \pm 2\pi/\lambda_o \tag{1.12}$$

being λ_o the wavelength in vacuum. The plus sign is for a forward wave and the minus sign is for a backward wave. If the wave is in a medium of refractive index n, it is $k = 2\pi/\lambda = 2\pi n/\lambda_o$.

We have further

$$\mathbf{E} = -(Z/k)(\mathbf{k} \times \mathbf{H}) \tag{1.13a}$$

$$\mathbf{H} = \mathbf{k} \times \mathbf{E}/kZ \tag{1.13b}$$

where Z is the so-called *medium impedance*

$$Z = (\mu/\varepsilon)^{1/2}. \tag{1.14}$$

The Poynting vector **S** gives the direction in which energy propagates

$$\mathbf{S} = (1/2)(\mathbf{E} \times \mathbf{H}) = (1/2)(\mathbf{k}/k)|E|^2/Z \tag{1.15}$$

and for plane waves propagating in an isotropic medium has the same direction of **k**.

The mean intensity is

$$I = |E^2|/2Z. \tag{1.16}$$

Plane waves, in a strict sense, do not exist. By definition, they should have infinite time-space extension. They are however, very useful and simplify many discussions; so they are our preferred kind of waves used in this book. We may use them as an approximation in limited space-time regions.

Spherical waves, or waves that have origin in a point, do instead exist. A harmonic spherical wave is written as

$$E = (Eo/r)e^{i(\omega t - kr)}. \tag{1.17}$$

A spherical wave has an amplitude which decreases with the distance r from its origin and its wavefront is a sphere. A plane wave can be seen as a portion of a spherical wave wavefront for r very large, in a region of dimensions « r.

1.4 Phase Velocity

A one dimensional sinusoidal wave propagating along x can be expressed as

$$\mathbf{E} = \mathbf{A}e^{i(\omega t - kx)}. \tag{1.18}$$

If we look at how its wavefront (constant phase surface) moves, we have

$$\varphi = \omega t - kx = \cos t \tag{1.19}$$

thus the phase velocity v_p is derived from

$$\Delta\varphi = \omega\Delta t - k\Delta x = 0, \tag{1.20}$$

or

$$\Delta x/\Delta t|_{\text{at const phase}} = \omega/k = v_p = \nu\lambda. \tag{1.21}$$

In vacuum $v_p = c$. In a medium of refractive index n, it is $v_p = c/n$.

Infinite plane or spherical waves do not carry information other than being or not there. Information is carried by finite wavepackets that consist of a group of sinusoidal waves with frequencies between ω and ω + dω. The velocity with which the envelope of the wavepacket moves is the velocity at which information can be transmitted and is called the group velocity v_g

$$v_g = d\omega/dk. \tag{1.22}$$

1.5 Dispersion

The velocity of a wave in a material medium is given by (1.7), where $(\varepsilon_r\mu_r)^{1/2} = n$ is the refractive index of the medium. Most materials used in optics are non-magnetic ($\mu_r = 1$), and therefore

$$n = (\varepsilon_r)^{1/2}. \tag{1.23}$$

Because ε_r varies with frequency, also n is a function of frequency or λ. This is the phenomenon of *dispersion*, and the media having this property are called *dispersive media*.

The classical theory of dispersion and absorption of radiation is due principally to P Drude (1863–1906)[2] and W Voigt (1850–1919)[3] and was elucidated fully by H A Lorentz (1853–1928).[4]

In the classical Lorentz model for dielectrics, the atom is considered to be formed by some oscillators whose frequencies are equal to the atomic absorption frequencies, ω_i. If such oscillators are modelled as particles with charge e and mass m, under the action of an oscillating electric field E, we may write for the position vector **r** of the generic oscillator with proper frequency (eigenfrequency) ω_o

$$\ddot{\mathbf{r}} + \gamma\dot{\mathbf{r}} + \omega_o^2\mathbf{r} = -(e/m)\mathbf{E}e^{i\omega t}, \tag{1.24}$$

where γ is a damping factor. A solution of equation (1.24) is

$$\mathbf{r} = -\left[(e/m)\mathbf{E}e^{i\omega t}\right]/\left[(\omega_o^2 - \omega^2) + i\omega\gamma\right], \tag{1.25}$$

[2]Drude P (1900) Ann Phys Lpz. 1(4):437

[3]Voigt W (1916) Magneto-Elektrooptik. G B Teubner, Leipzig

[4]Lorentz HA (1916) Theory of electrons. G B Teubner, Leipzig

and the oscillator has an electric dipole moment

$$\mathbf{p} = -e\mathbf{r}. \tag{1.26}$$

The polarization **P** produced by the vibrating electric field (assuming all oscillators to be equal and homogeneously distributed) is therefore written as

$$\mathbf{P} = -eN\mathbf{r} = N\alpha\mathbf{E}e^{i\omega t} = \varepsilon_o\chi\mathbf{E}e^{i\omega t} \tag{1.27}$$

where N is the number of oscillators (atoms) per unit volume and the *polarizability* $\alpha = \varepsilon_o\chi/N$ is a complex number (Drude's formula) that from (1.25) and (1.27) results

$$\alpha = \left[(e^2/m)\right] / \left[(\omega_o^2 - \omega^2) + i\omega\gamma\right]. \tag{1.28}$$

If ω is not very near to ω_o, the imaginary term in the denominator of (1.25) can be neglected. By introducing a complex refractive index ñ defined by

$$\tilde{\varepsilon}_r = \varepsilon_1 + i\varepsilon_2 = 1 + (P/\varepsilon_o E) = (\tilde{n})^2 = (n - ik)^2 \tag{1.29}$$

where $\tilde{\varepsilon}_r$ is the complex relative dielectric constant and n and k are the real and imaginary part of ñ respectively, we have therefore

$$(\tilde{n})^2 = (n^2 - k^2) - i(2nk) = 1 + N\alpha/\varepsilon_o. \tag{1.30}$$

By using equation (1.28) we have finally

$$\begin{aligned}(\tilde{n})^2 = (n^2 - k^2) - i(2nk) &= 1 + (N/\varepsilon_o)[\mathrm{Re}(\alpha) + \mathrm{Im}(\alpha)] \\ &= 1 + (e^2N/\varepsilon_o m)\{(\omega_o^2 - \omega^2)/[(\omega_o^2 - \omega^2)^2 + \omega^2\gamma^2]\} \\ &\quad - i(e^2N/\varepsilon_o m)\{\omega\gamma/[(\omega_o^2 - \omega^2)^2 + \omega^2\gamma^2]\}.\end{aligned} \tag{1.31}$$

In the case of a low density gas we can make the following approximations

$$k \ll 1, \quad |n - 1| \ll 1, \quad n^2 - 1 = 2(n - 1),$$

and obtain

$$n = 1 + (e^2N/\varepsilon_o m)\{(\omega_o^2 - \omega^2)/[(\omega_o^2 - \omega^2)^2 + \omega^2\gamma^2]\}, \tag{1.32}$$

and

$$k = (e^2N/\varepsilon_o m)\{\gamma\omega/[(\omega_o^2 - \omega^2)^2 + \omega^2\gamma^2]\}. \tag{1.33}$$

Making

$$k = \tilde{n}k_o = (n - ik)k_o \tag{1.34}$$

we obtain

$$E = Aexpi[\omega t - (n - ik)k_o x] = Aexp[-(kk_o x)]expi(\omega t - nk_o x) \tag{1.35}$$

which represents a wave with wavenumber *nko* propagating along x, whose amplitude decreases exponentially as the wave traverse is the medium. This exponential decrease of the amplitude and hence of the intensity of the light when traversing a medium is known as *Beer's law*. The absorption coefficient β is defined in terms of the field intensity

$$I(x) = I_o e^{-\beta x} \tag{1.36}$$

hence the absorption coefficient

$$\beta = 2k_o k \tag{1.37}$$

while n can be interpreted as the refractive index of the medium.

The behaviour of *k* and n near the resonant frequency is shown in Figs. 1.1a and b, respectively.

Far from resonance the refractive index n increases with increasing frequency, i.e. decreases with the wavelength. The corresponding negative wavelength coefficient dn/dλ is called *normal dispersion*.

The first observations of the dispersion phenomena were by Marcus in 1648 and Grimaldi (1613–1703) in 1665.[5]

Near resonance, however, the dispersion changes its sign; it is now called *anomalous dispersion* and in this region *k* and hence absorption become appreciable. In the ideal case of oscillation without damping ($\gamma = 0$) the refractive index n, instead of having a maximum and a minimum, tends towards $+\infty$ or $-\infty$ according to whether ν_o is approached from the lower or higher frequency side, respectively. Figure 1.1c shows the curve $\varepsilon = n^2$ for $\gamma = 0$. P Le Roux (1832–1907) was, in 1862, the first to observe anomalous dispersion.[6] In a medium the range where $\varepsilon < 0$ is characterised by total reflection. In this case, in fact, n would be purely imaginary and so would be the wave-vector k and therefore according to (1.35) a wave trying to enter the medium would have an exponential decreasing term with space.

Qualitatively (1.32) and (1.33) are in agreement with the measured profiles of the refractive index and absorption lines in gases. To have quantitative agreement,

[5]See Korff SA, Breit G (1932) Rev Mod Phys. 4:471

[6]Le Roux FP (1862) C R Acad Sci. Paris 40:126

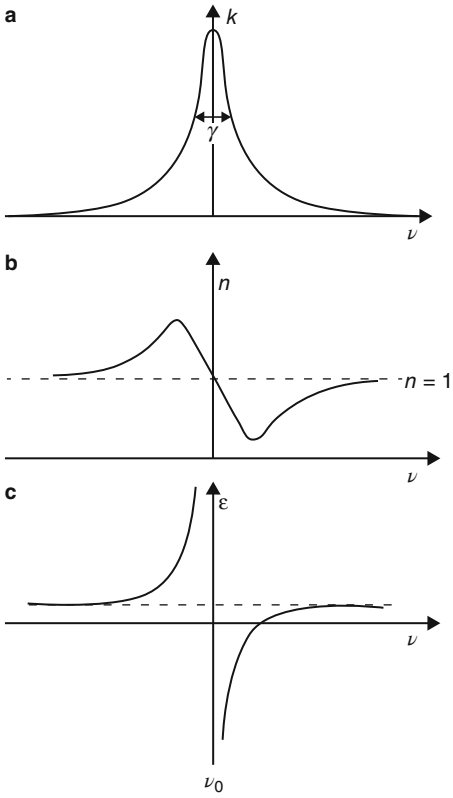


Fig. 1.1 Behaviour of k, n and ε as a function of frequency. (**a**) and (**b**) give k and n, respectively. In (**c**) the dielectric constant ε is given in the case $\gamma = 0$

however, the concept of effective number of oscillators has to be introduced. Equation (1.32) is accordingly written as[7]

$$n - 1 = (e^2/\varepsilon_o m)\sum_i N_i f_i(\omega_i^2 - \omega^2)/[(\omega_i^2 - \omega^2)^2 + \omega^2\gamma^2]. \tag{1.38}$$

N_i being the number among the N atoms whith eigenfrequencies ω_i. Note that the expression (1.38) for n assumes that all the atoms are in a single (ground) state.

The factor f_i was interpreted by Drude as the *number of dispersion electrons* per atom. W Pauli (1900–1958)[8] called it the *strength* of the oscillator. It expresses the probability for an atomic transition between two atomic levels separated by an energy $\hbar\omega_i$.

The classical equations considered above are in many cases in fairly good agreement with experiment and give a satisfactory interpretation of dispersion and also of absorption when the imaginary part of the refractive index is considered.

[7]Bertolotti M (2015) Masers and lasers: an historical approach. CRC Press, chapt 2

[8]Pauli W (1926) Quantentheorie, Handbuch der Physik, vol. 23. Springer, Berlin, p 87

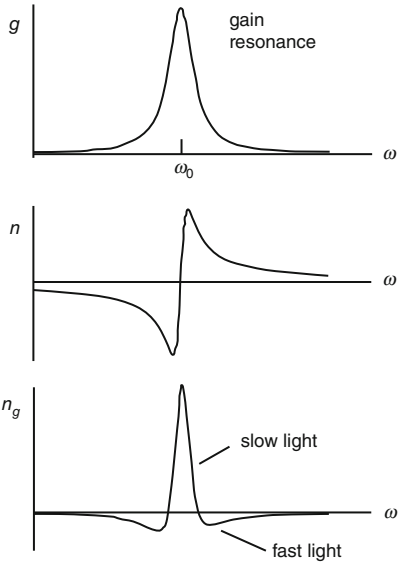


Fig. 1.2 The gain g (upper curve), real part of n (middle curve), group index n_g (lower curve) with the frequency ranges for slow and fast light in the case of an inverted population of atoms. (From Boyd[9] © Taylor & Francis)

The quantum mechanical version of (1.38) takes into account the stimulated emission. Accordingly in the case of a simple two level atom population, (1.38) is modified as

$$n - 1 = (e^2/m\varepsilon_o) f_{12}(N_1 - N_2)(\omega_{21}^2 - \omega^2)/[(\omega_{21}^2 - \omega^2)^2 + \omega^2\gamma^2], \tag{1.39}$$

where N_1 and N_2 are the population in the ground and excited level, respectively. The expression (1.39) shows that in the case of an inverted population $(N_1 - N_2) < 0$ the curves of Fig. 1.1 changes sign, as shown in Fig. 1.2. In this figure, the upper curve represents gain instead of attenuation and the curve representing n has reversed its behaviour, decreasing first and then increasing with increasing frequency.

1.6 Drude-Lorentz Model for Metals

The preceding model can be applied also to metals. In this case electrons are free and there is no restoring force; therefore $\omega_o = 0$. The metal is modelled as consisting of a free-electron gas with free-electron number density *N*, which moves against a fixed lattice background of positive ion cores. For alkali metals, this model holds for frequencies below the ultraviolet. For noble metals, interband transitions occurring at visible frequencies, limit its validity. In this plasma model, known as the *Drude model*, details of the lattice potential and electron-electron interactions are not taken into account. Instead, one simply assumes that some aspects of the band structure are incorporated into the effective optical mass m of each electron.

[9]Boyd RW (2009) J Mod Opt. 56:1908

Table 1.1 Plasma frequency for some metals $\hbar\omega_p$ in eV (from Rakic ref. 12)

Metal	$\hbar\omega_p$ (eV)	Plasma wavelength (nm)	m (effective mass)	τ (10^{-15} s)
Ag	9.01	137.64	0.96 ± 0.04	37.6
Au	9.03	137.34	0.99 ± 0.04	9.3 ± 0.9
Cu	10.83	114.51	1.49 ± 0.06	6.9 ± 0.7
Al	14.98	82.79		7.17
Be	18.51	67.00		4.50
Cr	10.75	115.36		
Ni	15.92	77.90	28	
Pd	9.72	127.59		
Pt	9.59	129.32	13	
Ti	7.29	170.12		
W	13.22	137.64		

Repeating calculations one finds a complex dielectric constant

$$\tilde{\varepsilon}_r(\omega) = 1 - \omega_p^2/(\omega^2 + i\gamma\omega) \tag{1.40}$$

where

$$\omega_p^2 = Ne^2/\varepsilon_o m, \tag{1.41}$$

which is the so-called *plasma frequency*. Table 1.1 gives ω_p for some metals, and the corresponding plasma wavelength in nm.

The real and imaginary components of the complex dielectric constant

$$\tilde{\varepsilon}_r(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega) \tag{1.42}$$

are given by

$$\varepsilon_1(\omega) = 1 - \omega_p^2\tau^2/(1 + \omega^2\tau^2) \tag{1.43a}$$

$$\varepsilon_2(\omega) = \omega_p^2\tau/\omega(1 + \omega^2\tau^2), \tag{1.43b}$$

where $\gamma = 1/\tau$, being $\tau \sim 10^{-14}$ s the *relaxation time of electrons*. For a gas of free electrons γ is a phenomenological damping constant, the relaxation time τ is a characteristic time for the distribution of electrons to approach equilibrium after a disturbance is removed.

Most applications in optics use frequencies well below the ultraviolet. If $\omega < \omega_p$ when the product $\omega\tau \gg 1$, ε is predominantly real and may be written as

$$\varepsilon(\omega) = 1 - \omega_p^2/\omega^2. \tag{1.44}$$

At low frequencies, where $\omega \ll \tau^{-1}$ it is $\varepsilon_2 \gg \varepsilon_1$, and the real and the imaginary part of the complex refractive index are of comparable magnitude with

$$n = k = (\varepsilon_2/2)^{1/2} = (\tau\omega_p^2/2\omega)^{1/2}. \tag{1.45}$$

In this region, metals are mainly absorbing, with an absorption coefficient β

$$\beta = (2\omega_p^2\tau\omega/c^2)^{1/2}. \tag{1.46}$$

The field fall off inside the metal as $e^{-z/\delta}$, where δ is the *skin depth*

$$\delta = 2/\beta = c/k\omega. \tag{1.47}$$

Figure 1.3 shows the real and imaginary part of ε for gold and Fig. 1.4 for silver.[10]

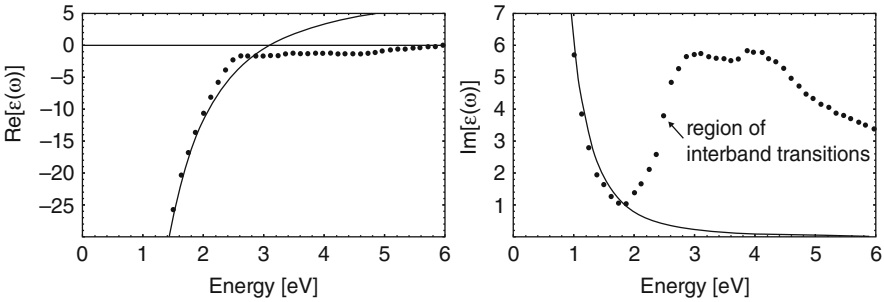


Fig. 1.3 Dielectric function $\varepsilon(\omega)$ eq. (1.43) of the free electron gas from Drude theory (solid line) and experimental values for gold (dots). Interband transitions limit the validity of this model at visible and higher frequencies. (From Maier[10] who used experimental data from Johnson and Christy[11])

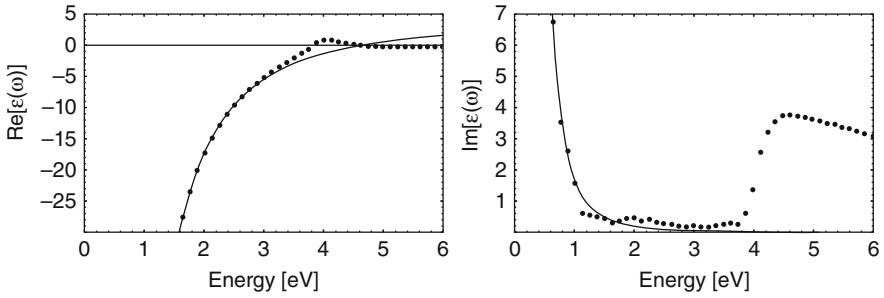


Fig. 1.4 The real and imaginary part of $\varepsilon(\omega)$ for silver determined by Johnson and Christy (dots) and the Drude model (solid line). (From Maier)[10]

[10]Maier SA (2007) Plasmonics; fundamentals and applications. Springer, New York

[11]Johnson PB, Christy RW (1972) Phys Rev. B6:4370

It may be seen that the dielectric function of the Drude model adequately describes the optical response of metals only for photon energies below the threshold of transition between electronic bands.

Above the band edge thresholds, photons are efficient in inducing interband transitions, raising electrons from the filled d band below the Fermi surface to higher bands. For practical purposes the Drude model can be integrated using (1.24) describing each interband transitions with the classical picture of a bound electron with resonance frequency ω_j. We may note that a number of equations of this form might have to be solved (each resulting in a separate contribution to the total polarization) in order to model the dielectric constant for noble metals accurately. Each interband transition leads to a Lorentz-oscillator term of the form $A_i/(\omega_i^2 - \omega^2 - i\gamma\omega)$ added to the free-electron result. One may write

$$\varepsilon(\omega) = \varepsilon_\infty - \sum_{j=1}^{q} f_j \omega_p^2 / [(\omega_j^2 - \omega^2) + i\gamma_j \omega] \tag{1.48}$$

where ε_∞ is a suitable constant, q is the number of oscillators with frequency ω_j, strength f_j, and lifetime $1/\gamma_j$.

In the case of metals absorption is obviously linked to conductivity σ. Defining the current density as

$$j = -Nev \tag{1.49}$$

where v is the drift velocity of electrons, and substituting into the equation of motion (1.24) with $\omega_o = 0$, we obtain

$$dj/dt + \gamma j = (Ne^2/m)E. \tag{1.50}$$

Using a local approximation to the current-field relation, we can assume that the temporal dependence of the electric field and current density are

$$E = E_o e^{-i\omega t} \quad \text{and} \quad j = j_o e^{-i\omega t}. \tag{1.51}$$

Replacing in (1.50) we have

$$j = (Ne^2\tau/m)[1/(1 - i\omega\tau)]E = \sigma E. \tag{1.52}$$

For a static field ($\omega = 0$), we have the static conductivity

$$\sigma_{st} = Ne^2\tau/m. \tag{1.53}$$

For the general case of an oscillating applied field, the conductivity becomes the dynamic conductivity, a complex quantity

$$\sigma = (Ne^2\tau/m)[1/(1 - i\omega\tau)]. \tag{1.54}$$

For very low frequencies ($\omega\tau \ll 1$), the dynamic conductivity is purely real and the electrons oscillate in phase with the electric field. As the frequency of the applied field increases, the inertia of the electrons introduces a phase lag and the dynamic conductivity becomes complex. For optical frequencies and above ($\omega\tau \gg 1$), the dynamic conductivity becomes purely imaginary and the electrons oscillate 90° out of phase with the electric field. Notice that the Drude conductivity of a material is proportional to its relaxation time.

The real and imaginary parts of the dynamic conductivity

$$\sigma = \sigma_1 + i\sigma_2 \tag{1.55}$$

are

$$\sigma_1 = (Ne^2\tau/m)\left[1/(1+\omega^2\tau^2)\right] \tag{1.56a}$$

$$\sigma_2 = (Ne^2\tau/m)\left[\omega\tau/(1+\omega^2\tau^2)\right]. \tag{1.56b}$$

Comparison with (1.43) allows to write

$$\varepsilon_1 = 1 - \sigma_1/\varepsilon_o\tau \tag{1.57a}$$

$$\varepsilon_2 = \sigma_2/\varepsilon_o\tau\omega. \tag{1.57b}$$

Other models can be also used. A more complete set of measurements for the metals listed in Table 1.1 has been made by Rakic et al.[12]

Similar considerations of the ones made with relation to (1.48) should be made for dielectrics. Also semiconductors present regions in which the real part of ε is negative. To shift the plasma frequency of semiconductors towards the blue region of the spectrum, one needs to heavily dope them, a practice that however, increases their losses. A quest for alternative materials to noble metals, has brought to a great variety of new materials, as Titanium nitride, or chalcogenides. Two dimensional plasmonic materials such as 2D electron gas systems or graphene are recently considered materials. *Topological insulator* is a material that behaves as an insulator in its interior but whose surface contains conducting states, meaning that electrons can move along the surface of the material. An example is $Bi_{1.5}Sb_{0.5}Te_{1.8}Se_{1.2}$ which has been recently studied by the Zheludev group.[13]

[12]Rakic AD et al (1998) Appl Opt. 37:5271

[13]Ou J-Y et al (2014) Nature Comm. 5:5139, doi: 10.1038/ncomms6139

1.7 Pulses and Group Velocity

An impulse of finite duration is a wave-packet with some frequency distribution. Its propagation in a linear medium may be described as a suitable superposition of plane waves of different frequency, because Maxwell's equations are linear as

$$\Psi(z,t) = \int A(k)\, e^{i(\omega t - kz)} dk \tag{1.58}$$

where A(k) gives the amplitude of the plane wave with wavenumber k. Considering the Ψ at t = 0, A(k) is the Fourier transform of Ψ(z,0) and $|A(k)|^2$ is the Fourier spectrum of Ψ(z,t). The relation between k and ω is the dispersion relation. In an isotropic medium the dispersion properties cannot depend on the direction of propagation and therefore

$$\omega(k) = \omega(-k). \tag{1.59}$$

An impulse is characterized by its central frequency ω_o and the frequency width $\Delta\omega$ around ω_o. That is A(k) is sharply peaked around k_o.

To study the time evolution of the pulse we may expand the frequency in the neighbourhood of k_o as a Taylor series

$$\omega(k) = \omega_o + (d\omega/dk)_o(k - k_o) + (1/2)(d^2\omega/dk^2)_o(k-k_o)^2 + \cdots \tag{1.60}$$

Considering only the first term of the expansion in (1.58) we have

$$\Psi(z,t) = \exp i(\omega_o t - kz) \int A(k) \exp i\left[(d\omega/dk)_o t - z\right](k - k_o) dk. \tag{1.61}$$

The integral is the Fourier transform of a function of the kind

$$E\left[z - (d\omega/dk)_o t\right]. \tag{1.62}$$

Thus, the amplitude of the pulse can be written as

$$\Psi(z,t) = \exp i(\omega_o t - kz) E\left[z - (d\omega/dk)_o t\right]. \tag{1.63}$$

Equation (1.63) represents a signal carried by a carrier-frequency ω_o and characterized by a modulation curve $E[z - (d\omega/dk)_o t]$. This modulation curve moves with a velocity

$$v_g = dz/dt = d\omega/dk, \tag{1.64}$$

that is the *group velocity*.

In a medium where the phase velocity is a constant and does not depend upon frequency

$$v_p = v_g. \tag{1.65}$$

This approximation is legitimate only when the distribution A(k) is sharply peaked at k_o and the frequency ω is a smoothly varying function of k around k_o. If the electromagnetic energy density of the pulse is associated with the absolute square of the amplitude, the group velocity represents the transport of energy.

Using the refractive index n, the phase velocity is

$$v_p = c/n(\omega), \tag{1.66}$$

and is greater or smaller than c depending on whether n(ω) is smaller or larger than unity. The relation between k and ω is

$$k = n(\omega)\omega/c. \tag{1.67}$$

Therefore, the group velocity from (1.64) is

$$v_g = c/n_g = c/[n + \omega(dn/d\omega)], \tag{1.68}$$

where n_g is the group index

$$n_g = n + \omega(dn/d\omega). \tag{1.69}$$

For normal dispersion, $dn/d\omega > 0$ and the group velocity is less than the phase velocity. Near an absorption line, $dn/d\omega$ may be either positive or negative and very large. In the first case, the group velocity is very small and one speaks of *slow light*. In regions of anomalous dispersion, on the contrary, $dn/d\omega$ can become large and negative. Then the group velocity may become larger than c and one speaks of *fast light*. This occurs when $dn/d\omega$ is very negative, which is equivalent to a rapid variation of ω as a function of k, thus making the approximation (1.60) illegitimate. Therefore there is no violation of the relativity theory.

To see the role played by the absorption, Fig. 1.5 allows to see why slow light is expected in the wing of an absorption line and fast light is expected near line centre. The reverse is true in the case of an inverted medium as shown in the lower curve of Fig. 1.2.

The problem of slow and fast light has been considered by various authors.[14]

In a dispersive medium each component travels with a different velocity and therefore changes its phase. A laser pulse propagating in a dispersive medium will

[14]See for example Milonni PW (2005) Fast light, slow light and left-handed light. IoP, Bristol

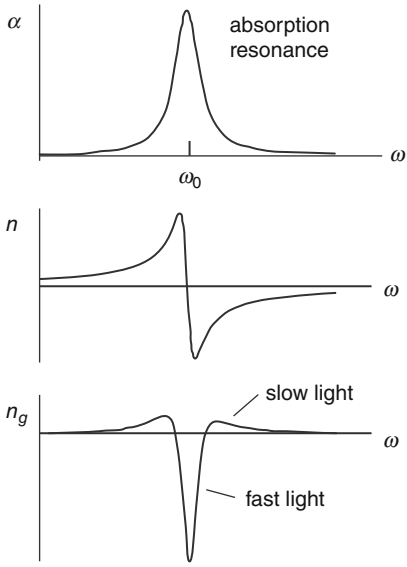


Fig. 1.5 Absorption α, real part of n and group index n_g with the frequency regions for slow and fast light. (From Boyd © Taylor & Francis)

remain undistorted in shape provided the term $(d^2\omega/dk)(k - ko)^2$ in (1.60) can be neglected. If this is not the case, the pulse spreads when travelling. The spreading of the pulse can be accounted for assuming the group velocity spreads

$$\Delta v_g = (d^2\omega/dk^2)\Delta k. \tag{1.70}$$

1.8 Polarization

The direction in which the wave electric field vibrates is called *polarization direction*. If we consider a sinusoidal wave we may distinguish the following cases:

- if while travelling the electric field of the wave oscillates always along the same direction, the wave is said to be *linearly polarized*. If this direction is for example the x direction and the wave travels in the z direction, the electric field oscillates in the xz plane that is called *polarization plane*.
- If the tip of the arrow representing the electric field does no more oscillate in the same direction but describes a circular motion with constant amplitude around the propagation direction, than the wave is said to be *circularly polarized*. If the electric vector, observed looking toward the source, rotates in clockwise way, the light is said to be *clockwise circularly polarized* or *right-handed circularly polarized*. If it rotates in the other direction the light is said to be *anticlockwise* (or *left-handed*) circularly polarised.
- If the tip of the arrow representing the electric field describes an ellipse the light is said to be (clockwise or anticlockwise) *elliptically polarized*.

– If the vector of the electric field oscillates randomly the wave is said to be *unpolarized*.
– A wave may also be partially polarized.

Our definition of polarization refers to the behaviour of the electric field. It is also possible to refer to the magnetic field, therefore it is advisable always specify to which vector one considers.

To describe the behaviour of polarization we observe that the electric field of a plane wave propagating in the z-direction, in the plane xy can be obtained as the sum of two components, one along x and the other along y, oscillating at the same frequency ω with possibly a phase difference δ between them

$$E_x = A_x \cos(\omega t - kz) \quad \text{and} \quad E_y = A_y \cos(\omega t - kz + \delta). \tag{1.71}$$

The curve described by the tip of the electric field is completely analogous to the motion of a bi-dimensional harmonic oscillator. We derive the curve eliminating the quantity ωt − kz in (1.71)

$$\begin{aligned}(E_y/A_y)^2 &= \cos^2(\omega t - kz)\cos^2\delta + \sin^2(\omega t - kz)\sin^2\delta \\ &\quad - 2\cos(\omega t - kz)\sin(\omega t - kz)\cos\delta\sin\delta \\ &= (E_x/A_x)^2\cos^2\delta + \left[1 - (E_x/A_x)^2\right]\sin^2\delta - 2(E_x/A_x) \\ &\quad \times \cos\delta\left[-(E_y/A_y) + (E_x/A_x)\cos\delta\right]\end{aligned} \tag{1.72}$$

because

$$\sin(\omega t - kz)\sin\delta = -(E_y/A_y) + \cos(\omega t - kz)\cos\delta. \tag{1.73}$$

After some algebra we find

$$\begin{aligned}(E_y/A_y)^2 &= (E_x/A_x)^2(\cos^2\delta - \sin^2\delta - 2\cos^2\delta) + 2(E_x/A_x)(E_y/A_y)\cos\delta + \sin^2\delta \\ &= -(E_x/A_x)^2 + 2(E_x/A_x)(E_y/A_y)\cos\delta + \sin^2\delta.\end{aligned} \tag{1.74}$$

That is

$$(E_x/A_x)^2 + (E_y/A_y)^2 - 2(E_x/A_x)(E_y/A_y)\cos\delta = \sin^2\delta. \tag{1.75}$$

This is the equation of an ellipse which is inscribed into a rectangle with its sides parallel to the axes and of length $2A_x$ and $2A_y$, respectively. The ellipse is tangent to the rectangle sides in the points (Fig. 1.6)

$$(\pm A_x,\ \pm A_y\cos\delta) \quad \text{and} \quad (\pm A_x\cos\delta,\ \pm A_y). \tag{1.76}$$

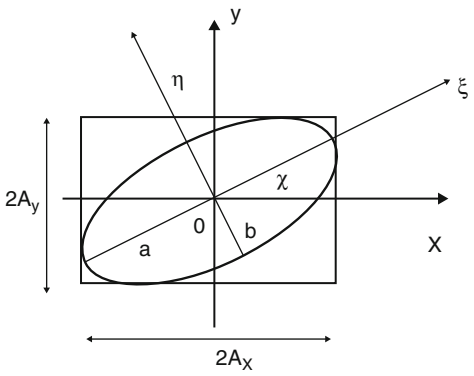


Fig. 1.6 The ellipse describing the behaviour of the electric field components

In general, the ellipse axes ξ and η do not coincide with x and y. The principal semi-axes of the ellipse a,b and the inclination χ of the major axis, are

$$\begin{aligned} a^2 + b^2 &= A_x^2 + A_y^2 \\ \mathrm{tg}2\chi &= \cos\delta \mathrm{tg}2\beta \end{aligned} \tag{1.77}$$

with

$$\mathrm{tg}\beta = (A_y/A_x) = [2A_xA_y/(A_x^2 - A_y^2)]\cos\delta. \tag{1.78}$$

The ellipse reduces to a straight line when

$$\delta = m\pi \quad (m = 0, 1) \tag{1.79}$$

In this case

$$(E_y/E_x) = (-1)^m(A_y/A_x), \tag{1.80}$$

and the light is linearly polarized.

If, on the contrary

$$\delta = \pm(1/2)\pi \quad \text{and} \quad A_y = A_x \tag{1.81}$$

the ellipse reduces to circle. If $\delta = -\pi/2$ the light is clockwise circularly polarized. If $\delta = +\pi/2$ the light is anticlockwise circularly polarized.

Figure 1.7 gives examples of some possible polarizations, in the general case that $A_x \neq A_y$ for some selected values of δ.

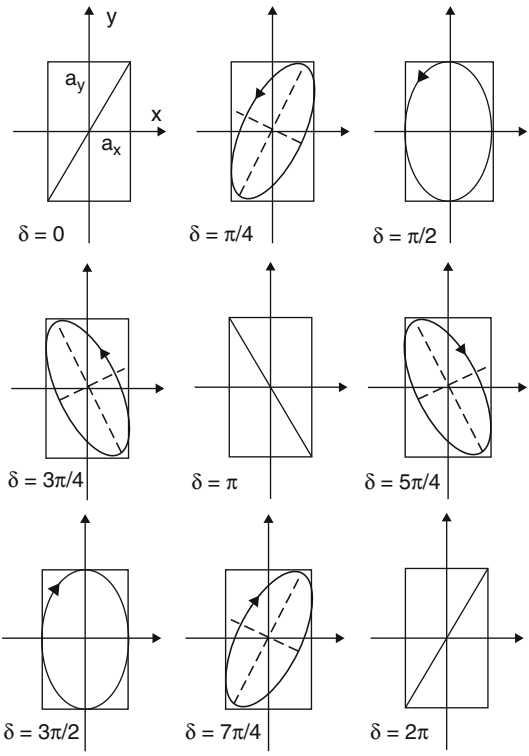


Fig. 1.7 Some examples of polarizations

1.9 Jones Matrices, Stokes Parameters and the Poincarè Sphere

The state of polarization of a field **E** propagating in the z-direction can be represented giving the two components x and y in the form of a two row matrix.

Jones in 1941[15] introduced the matrix

$$\mathrm{J} = \begin{vmatrix} A_x e^{i\delta x} \\ A_y e^{i\delta y} \end{vmatrix} \tag{1.82}$$

to describe the two field components. Whatever the state of polarization, the light field of a completely polarized beam can be resolved into x and y components that are coherent and in any z plane the amplitudes and phases of these components

[15]Jones RC (1941) J Opt Soc America 31:488; Jones RC (1941) J Opt Soc America 32:486 (1942). See also, Longhurst RS (1981) Geometrical and physical optics, 3rd edn. Longman, London and Hodgson N, Weber H (1997) Optical resonators, fundamentals, advanced concepts and applications. Springer

can be specified by the corresponding complex amplitudes. The matrix notation introduced by Jones was intended to allow calculation of the polarization changes suffered by a field crossing an arbitrary number of optical elements like retarders, polarizers, etc. In the most general example of an ideal system of polarizers and retarders each component of the incident light gives a contribution to each component of the emerging light, and since the contributions are still coherent they simply interfere to give the total x and y components and hence the intensity and state of polarization of the emergent light. Suppose, then, that the system is such that an incident x component of unit amplitude and zero phase gives emergent x and y components whose complex amplitudes are

$$J_{xx} = T_{xx} e^{i\delta xx}; \quad J_{yx} = T_{yx} e^{i\delta yx} \tag{1.83}$$

and that an incident y component of unit amplitude and zero phase gives an emergent x and y components whose amplitudes are

$$J_{xy} = T_{xy} e^{i\delta xy}; \quad J_{yy} = T_{yy} e^{i\delta yy}. \tag{1.84}$$

If the x and y components of the incident light have complex amplitudes ψ_x and ψ_y the complex amplitudes of the x and y component of the emergent light are given by

$$\begin{aligned} \psi_x^{'} &= J_{xx}\psi_x + J_{xy}\psi_y \\ \psi_y^{'} &= J_{yx}\psi_x + J_{yy}\psi_y \end{aligned} \tag{1.85}$$

or, in matrix form

$$\begin{vmatrix} \psi_x^{'} \\ \psi_y^{'} \end{vmatrix} = \begin{vmatrix} J_{xx} & J_{xy} \\ J_{yx} & J_{yy} \end{vmatrix} \begin{vmatrix} \psi_x \\ \psi_y \end{vmatrix}. \tag{1.86}$$

In a compact way (1.86) may be written as

$$\Psi' = J\Psi. \tag{1.87}$$

The column vectors Ψ, Ψ', which specify the incident and emerging fields respectively, are the Jones vectors and the matrix J corresponds to the Jones matrix of the system.

If light passes though a succession of n polarising and retarding components the Jones matrices of the components being $J_1, \ldots J_n$, one has

$$\begin{aligned} \Psi_1^{'} &= J_1\Psi_1 \\ \Psi_2^{'} &= J_2\Psi_2 = J_2\Psi_1^{'} = J_2J_1\Psi_1, \text{etc.} \end{aligned} \tag{1.88}$$

i.e.

$$\Psi_n' = J_{tot}\Psi_1 \tag{1.89}$$

where the total Jones matrix is given by

$$J_{tot} = J_n J_{n-1} \ldots J_2 J_1 \tag{1.90}$$

and must, of course, be evaluated by taking the product in this order.

A few examples may help in understanding the use of the matrices. The Jones vector

$$J = \begin{vmatrix} \cos\psi \\ \sin\psi \end{vmatrix} \tag{1.91}$$

describes a light linearly polarized with the electric field oscillating in a direction which forms an angle ψ with the x axis;

$$J = \begin{vmatrix} -\sin\psi \\ \cos\psi \end{vmatrix} \tag{1.92}$$

describes a light linearly polarized with the electric field oscillating in a direction which forms an angle $\psi + \pi/2$ with the x axis;

$$J = \begin{vmatrix} 1 \\ 0 \end{vmatrix} \tag{1.93a}$$

describes a light linearly polarized with the electric field oscillating along the x axis;

$$J = \begin{vmatrix} 0 \\ 1 \end{vmatrix} \tag{1.93b}$$

describes a light linearly polarized with the electric field oscillating along the y axis;

$$R = \left(1/\sqrt{2}\right) \begin{vmatrix} 1 \\ -i \end{vmatrix} \tag{1.93c}$$

describes a light clockwise circularly polarized and

$$L = \left(1/\sqrt{2}\right) \begin{vmatrix} 1 \\ i \end{vmatrix} \tag{1.93d}$$

describes a light anticlockwise circularly polarized.

To characterize the polarization ellipse, three independent quantities are necessary, the amplitudes A_1 and A_2 and the phase difference δ, or the major and minor axes a, b and the angle χ which specifies the orientation of the ellipse. Another way to characterise polarization is then to use the Stokes parameters (Stokes[16] 1852). The Stokes parameters for a plane monochromatic wave are the four quantities

$$\begin{aligned} S_o &= A_1^2 + A_2^2 \\ S_1 &= A_1^2 - A_2^2 \\ S_2 &= 2A_1A_2\cos\delta \\ S_3 &= 2A_1A_2\sin\delta. \end{aligned} \tag{1.94}$$

Only three of them are independent since they are related by the identity

$$S_o^2 = S_1^2 + S_2^2 + S_3^2. \tag{1.95}$$

The parameter S_o is evidently proportional to the intensity of the wave. The parameters S_1, S_2 and S_3 are related in a simple way to the angle $\psi(0 < \psi < \pi)$ which specifies the orientation of the ellipse and the angle $\chi(-\pi/4 < \chi < \pi/4)$ which characterizes the ellipticity and the sense in which the ellipse is being described. In fact the following relations hold

$$S_1 = S_o\cos 2\chi\cos 2\psi \tag{1.96a}$$

$$S_2 = S_o\cos 2\chi\sin 2\psi \tag{1.96b}$$

$$S_3 = S_o\sin 2\chi. \tag{1.96c}$$

These relations indicate a simple geometrical representation of all different states of polarization: S_1, S_2 and S_3 may be regarded as the Cartesian coordinates of a point P on a sphere Σ of radius S_0 such that 2χ and 2ψ are the spherical angular coordinates of this point (Fig. 1.8).

This sphere is called *Poincarè sphere*.[17] Thus to every possible state of polarization of a plane monochromatic wave of a given intensity (S_o = constant), there corresponds one point on Σ and vice versa. Since χ is positive or negative according as the polarization is right-handed or left-handed, it follows from (1.96c) that right-handed polarization is represented by points on Σ which lie above the equatorial plane (xy-plane), and left-handed polarization by points on Σ which lie below this plane. Further, for linearly polarized light the phase difference δ is zero or an integral multiple of π; according to (1.94) the Stokes parameter S_3 is then

[16]Stokes GG (1852) Trans Cambr Phil Soc. 9:399, reprinted in his Mathematical and physical papers, vol. III. Cambridge Univ. Press, p. 233, 1901

[17]Poincarè H (1892), Theorie Mathematique de la Lumiere, vol. 2. George Carrè, Paris, chap. 12

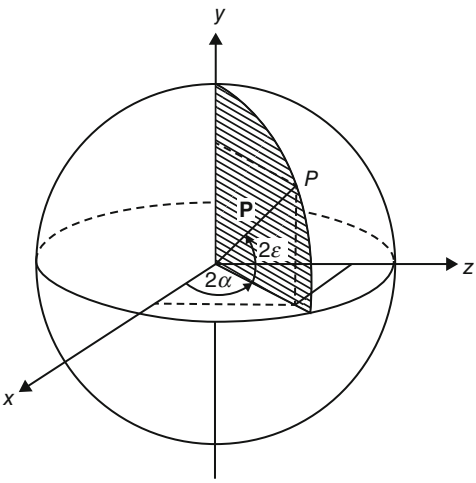


Fig. 1.8 Poincarè sphere to represent polarization states

zero, so that linear polarization is represented by points in the equatorial plane. For circular polarization $A_1 = A_2$ and $\delta = \pi/2$ or $-\pi/2$ according as the polarization is right- or left-handed; hence right-handed circular polarization is represented by the north pole ($S_1 = S_2 = 0$, $S_3 = S_o$) and left-handed circular polarization by the south pole ($S_1 = S_2 = 0$, $S_3 = -S_o$).

1.10 Optically Anisotropic Media

An optically anisotropic medium is characterized by a dielectric tensor

$$\begin{vmatrix} D_x \\ D_y \\ D_z \end{vmatrix} = \begin{vmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{vmatrix} \begin{vmatrix} E_x \\ E_y \\ E_z \end{vmatrix} \tag{1.97}$$

which is symmetric, that is

$$\varepsilon_{ij} = \varepsilon_{ji}. \tag{1.98}$$

Moreover, it is possible to rotate the axes to find a system of axes (*principal axes*) for which

$$\begin{vmatrix} D_x \\ D_y \\ D_z \end{vmatrix} = \begin{vmatrix} \varepsilon_x & 0 & 0 \\ 0 & \varepsilon_y & 0 \\ 0 & 0 & \varepsilon_z \end{vmatrix} \begin{vmatrix} E_x \\ E_y \\ E_z \end{vmatrix} = \varepsilon_o \begin{vmatrix} n_x^2 & 0 & 0 \\ 0 & n_y^2 & 0 \\ 0 & 0 & n_z^2 \end{vmatrix} \begin{vmatrix} E_x \\ E_y \\ E_z \end{vmatrix} \tag{1.99}$$

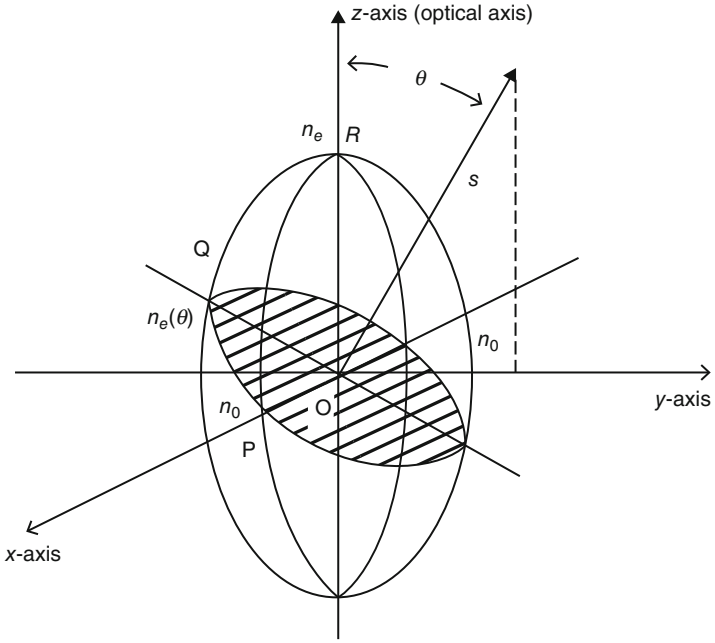


Fig. 1.9 The ellipsoid of indices

A plane wave which propagates along any direction can be decomposed into two parts that vibrate on orthogonal directions and find two refractive indices n_1 and n_2 which correspond to two different phase and group velocities.

The displacement vectors $\mathbf{D_1}$ and $\mathbf{D_2}$ are orthogonal between them. To obtain n_1 and n_2, $\mathbf{D_1}$, $\mathbf{D_2}$, $\mathbf{E_1}$, $\mathbf{E_2}$, $\mathbf{H_1}$, and $\mathbf{H_2}$ the ellipsoid of indices is usually used.

The index ellipsoid is shown in Fig. 1.9 and is represented by the following expression

$$x^2/\varepsilon_{rx} + y^2/\varepsilon_{ry} + z^2/\varepsilon_{rz} = 1 = x^2/n_x^2 + y^2/n_y^2 + z^2/n_z^2. \tag{1.100}$$

To find n_1 and n_2 for a given direction of propagation $\mathbf{s}$ one needs to draw the plane passing through the ellipsoid origin and perpendicular to $\mathbf{s}$. The intercept of this plane with the index ellipsoid gives an ellipse. The two principal axes of the ellipse correspond to $2n_1$ and $2n_2$, respectively.

The corresponding $\mathbf{D_1}$ and $\mathbf{D_2}$ are parallel to the principal axes of the ellipse. The vector $\mathbf{E}$ is obtained by

$$\mathbf{D} = n^2 \varepsilon o[\mathbf{E} - \mathbf{s}(\mathbf{s} \cdot \mathbf{E})] \tag{1.101}$$

where $\mathbf{s}$ is the unit vector in the s direction.

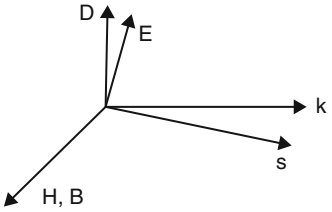


Fig. 1.10 The disposition of the electromagnetic vectors

This relation is easily obtained by Maxwell equations. Substituting

$$D = D_o e^{i(\omega t - nk \cdot r)} \tag{1.102}$$

into (1.1a) with $\rho_{ext} = 0$ we have

$$\text{div}\mathbf{D} = -in\mathbf{k} \cdot \mathbf{D} = 0. \tag{1.103}$$

For H one obtains, in a similar way using (1.1b)

$$\text{div}\mathbf{H} = in\mathbf{k} \cdot \mathbf{H} = 0. \tag{1.104}$$

The two relations show that **D** and **H** should be orthogonal to **k**. By inserting (1.10a) and (1.103) into (1.1d) with $\mathbf{j_{ext}} = 0$, we have

$$in\mathbf{kxH} = -i\omega\mathbf{D} \tag{1.105}$$

so that **D, H** and **k** make the tern shown in Fig. 1.10.

By using (1.1c) we have similarly

$$in\mathbf{kxE} = i\omega\mu_o\mathbf{H} \tag{1.106}$$

so that **H** must be orthogonal to the plane (**E,k**) which means that **E** should stay in the plane (**D,k**). The Poynting vector **S** = **ExH** lies also in the plane (**D,k**). If the medium were electrically isotropic **D** and **E** would be parallel to each other and so would do **k** and **S**. In an anisotropic medium **D** and **E** make an angle between them and the propagation direction of the wave and the propagation direction of the energy are not in the same direction.

From (1.105) we have for the modulus of **H**

$$H = \omega D/nk. \tag{1.107}$$

If $E_\perp$ is the **E** component orthogonal to **k**, from (1.106) one has

$$E_\perp = \omega\mu_o H/nk = \omega\mu_o D/n^2k^2 = D/\varepsilon_o n^2 \tag{1.108}$$

taking into account (1,10a). If **s** is the unit vector in the direction of **k** we may write

$$E - (\mathbf{s} \cdot \mathbf{E}) = D/\varepsilon_o n^2 = E_\perp \tag{1.109}$$

where $\mathbf{s} \cdot \mathbf{E}$ is the E component parallel to k and we have used (1.108). Equation (1.109) gives immediately (1.101).

The determination of the polarization may be obtained through the following considerations. If, for example, we have a wave propagating in the z direction with the electric vector parallel to the x-direction, it induces only a polarization

$$P_x = \varepsilon_o \chi_{11} E_x = D_x - \varepsilon_o E_x = (\varepsilon_{11} - 1)\varepsilon_o E_x \tag{1.110}$$

and finds a dielectric constant ε_{11}, that is a refractive index n_x.

If on the contrary the wave is polarized with the electric field oscillating along the y-direction, it finds a refractive index n_3.

Each un-polarized wave propagating in the z-direction may be decomposed into two components with polarizations parallel to the x and y directions, respectively. These two components travel in the crystal with different velocities.

Uniaxial crystals are characterized by a *principal axis* (conventionally the z-axis) along with the dielectric constant is ε_{33}. Perpendicularly to this axis the dielectric constant does not depend on the direction, that is $\varepsilon_{11} = \varepsilon_{22}$. The z-direction is unique because the propagation velocity is independent from the polarization and is called *optical axis*. The difference between n_x (=n_y) and n_z is a measure of birefringence that is said to be positive or negative according to $n_z - n_x$ being larger or lower than 0 respectively.

Crystals with less symmetry have two optical axes and are called bi-axial. They have $n_x \neq n_y \neq n_z$.

1.10.1 Uniaxial Crystals

Uniaxial crystals have an index ellipsoid with circular symmetry around z (the optical axis) as shown in Fig. 1.9. The index $n_x = n_y$ is said to be the *ordinary index* n_o and n_z is called *extraordinary index* n_e. Accordingly the index ellipsoid is written as

$$x^2/n_o^2 + y^2/n_o^2 + z^2/n_e^2 = 1. \tag{1.111}$$

If a light ray propagates in a direction s which makes an angle θ with the optical axis (z), due to the circular symmetry we may chose the y axis to coincide with the projection of s onto the xy plane. The plane orthogonal to r intercept the ellipsoid into an ellipse (shaded in the figure)

$$y^2/n_o^2 + z^2/n_e^2 = 1. \tag{1.112}$$

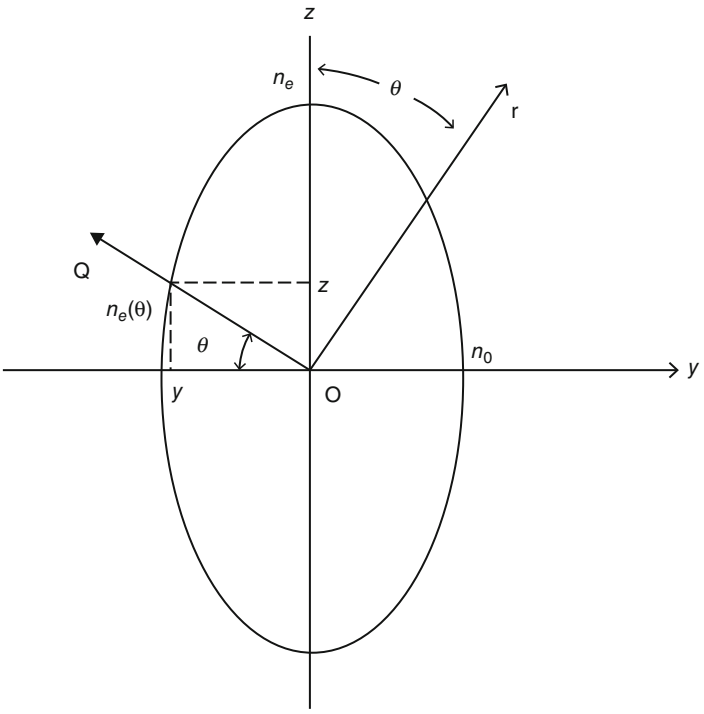


Fig. 1.11 The plane section of the index ellipsoid for the calculation of n_e

The two allowed polarization directions are parallel to the axes of this ellipse and correspond to OP and OQ (fig. 1.9).

The two waves polarized along these directions have refractive indices given by OP = n_o and OQ = $n_e(\theta)$. In the case of the extraordinary wave, the polarization plane changes with θ and so does the refractive index. We make reference to Fig. 1.11 to write

$$n_e^2(\theta) = z^2 + y^2 \quad \text{and} \quad z = n_e(\theta)\sin\theta. \tag{1.113}$$

By substituting (1.113) into the equation of the ellipse (1.102), one has

$$1/n_e^2(\theta) = \cos^2\theta/n_o^2 + \sin^2\theta/n_e^2. \tag{1.114}$$

Therefore for θ = 0°, that is for a propagation along the optical axis $n_e(0°) = n_o$, while for θ = 90°, $n_e(90°) = n_e$. For propagation parallel to the optical axis (z-direction) the birefringence is absent because the orthogonal section of the ellipsoid to the direction of z is a circle.

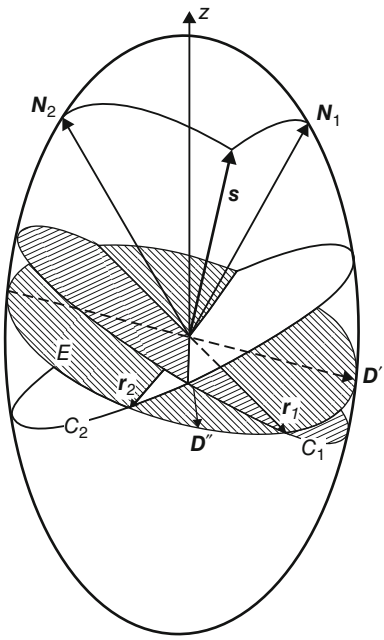


Fig. 1.12 A construction for determining the planes of vibration (s,D′) and (s,D″). Born and Wolf[18]

1.10.2 Biaxial Crystals

In this case the direction of vibration of a wave through the crystal can be found with the same operation shown in Fig. 1.9. Because the three refractive indices are all different from each other, however, the plane normal to the direction of propagation will cut the ellipsoid forming an ellipse and not a circle as in the case of uniaxial crystals. The direction of vibrations of the two vectors **D** corresponding to a given direction of propagation **s** are directed along the two principal axes of the ellipse.

There is also another construction by which the directions of vibrations can be determined. It is known that an ellipsoid has two circular sections C_1 and C_2 passing through the centre, and that the normals $\mathbf{N_1}$ and $\mathbf{N_2}$ to these sections are coplanar with the longest and shortest principal axes of the ellipsoid. These two directions $\mathbf{N_1}$ and $\mathbf{N_2}$ are called the optical axes. Since the sections C_1 and C_2 are circular (also of the same radius), the directions $\mathbf{N_1}$, $\mathbf{N_2}$ have the property that there is only one velocity of propagation along them: D can than take any direction perpendicular to s. Let E (Fig. 1.12) be the elliptical section through the centre, at right angle to an arbitrary unit normal s. This plane intersects the circles C_1, C_2 in two radial vectors r_1, r_2 which are of equal length and must therefore make equal angles with the principal axes of E. The required directions of vibration are therefore the bisectors of the directions r_1, r_2. But r_1 is perpendicular to N_1 and s,

[18]Born M, Wolf E (1964) Principles of optics, Oxford, p 675

and is, therefore, perpendicular to the plane containing N_1 and s; similarly r_2 is perpendicular to the plane containing N_2 and s. If these planes intersect the ellipse E in vectors r_1', r_2' the principal axes of the ellipse must also bisect the directions r_1', r_2'. It follows that the planes of vibrations of the electric displacement, i.e. the planes containing s and D′ or D″, are the internal or external bisectors of the angle between the planes (N_1,s) and (N_2,s).

1.11 Chirality

Chirality,[19] is the lack of reflection or inversion symmetry of an object that does not coincide with its mirror image and can occur in two forms (*enantiometers*) that are mirror images one of each other. It arises from different issues. Although often associated with biochemistry due to the numerous chiral bio-molecules, recently chirality has been obtained in engineered materials and has attracted much attention in relation of media that can be used for optical signal processing devices. From the optical point of view, chiral structures possess the ability to rotate the plane of polarization of electromagnetic waves (*optical activity*), and exhibit *dichroism*, that is the property to split a beam of polychromatic light into two beams with different wavelengths.[20] Therefore, a dichroic material can be considered either one which causes visible light to be split up into distinct beams of different wavelengths (*dichroic mirrors*) or one in which light rays having different polarizations are absorbed by different amounts. When the polarization states in question are right and left-handed circular polarization, one refers to *circular dichroism*.

In isotropic solutions of chiral molecules, optical activity effects arise from magnetic contributions to the linear optical response and are usually relatively weak. To describe the chiral response of a material it is convenient to write the constitutive relations between E, D, B and H using the temporal Fourier transforms, so that the frequency-domain constitutive relations can be expressed as

$$\begin{aligned} D(\omega) &= \varepsilon(\omega)E(\omega) - i\xi(\varepsilon_o\mu_o)^{1/2}H(\omega) \\ B(\omega) &= i\xi(\varepsilon_o\mu_o)^{1/2}E(\omega) + \mu(\omega)H(\omega) \end{aligned} \tag{1.115}$$

where ξ is the chirality parameter. It is dimensionless and is real in a lossless medium. For a homogeneous, isotropic chiral medium the space dependence of μ, ε and ξ is missing.[21]

[19]Lekner J (1996) Pure and Appl Optics 5: 417; Van Bladel JG (2007) Electromagnetic fields. Wiley; Valev VK (2016), Opt Phot. p 35; Valev VK(2013) Adv Mater. 25:2517

[20]Barron LD (2004) Molecular light scattering and optical activity. Cambridge Univ. Press, Cambridge.

[21]Lakhtakia A (1994) Beltrami fields in chiral media. World Scientific, Singapore; Mackay TG, Lakhtakia A (2008) Progress in Optics 51:121.

Assuming propagation along the z axis one finds two solutions of the Maxwell equations, in the form of a right and a left circular polarization state:

$$\begin{aligned}\mathbf{E} &= A(\mathbf{u_x} - i\mathbf{u_y})\exp[-ik_o(n-\xi)z] \\ \mathbf{H} &= i[(\varepsilon/\mu)^{1/2}A(\mathbf{u_x} - i\mathbf{u_y})]\exp[-ik_o(n-\xi)z]\end{aligned} \tag{1.116}$$

that is a right circular polarized wave and

$$\begin{aligned}\mathbf{E} &= C(\mathbf{u_x} + i\mathbf{u_y})\exp[-ik_o(n+\xi)z] \\ \mathbf{H} &= i[(\varepsilon/\mu)^{1/2}C(\mathbf{u_x} + i\mathbf{u_y})]\exp[-ik_o(n+\xi)z]\end{aligned} \tag{1.117}$$

that is a left circular polarized wave, being A and C two suitable constants.

These fields propagate with wave-numbers

$$\begin{aligned}k_1 &= \omega\{[(\varepsilon(\omega)\mu(\omega)]^{1/2} - \xi\} \\ k_2 &= \omega\{[(\varepsilon(\omega)\mu(\omega)]^{1/2} + \xi\}\end{aligned}. \tag{1.118}$$

Since $k_1 \neq k_2$, the left and right circularly polarized plane waves propagate with different phase velocities and with different attenuation rates, so the material is circularly birefringent.

To be chiral, a body must not contain any rotation-reflection axis. In other words, it may have neither symmetry planes, nor any center of inversion, nor any higher rotation-reflection symmetry. For example, a helix is a chiral object.

It has been predicted and demonstrated that strong enough optical activity may result in negative refraction and reflection, and this has further fostered research in new chiral dichroic schemes to be used as base elements in metamaterial constitution. Strong circular dichroism from metamaterials has been studied.[22]

Apart from 3D chiral objects, the possibility to obtain circular dichroism, i.e. optical activity, with non-chiral elements was studied in the past,[23] and recently reconsidered.[24] This phenomenon is obtained when the experimental configuration composed by both the nanopatterned surface and the optical incident field is non-superimposable on its mirror image. In this respect it is important to note that more than a metamaterial we are considering a "metasurface," with optical response strongly dependent on the input beam direction and polarization.[25]

[22]Decker M, Klein MW, Wegener M, Linden S (2007) Opt Lett. 32:856; Wegener M, Linden S (2009) Physics 2:3

[23]Williams R (1968) Phys Rev Lett. 21:342

[24]Volkov SN, Dolgaleva K, Boyd RW, Jefimovs K, Turunen J, Svirko Y, Canfield BK, Kauranen M (2009) Phys Rev. A79:043819

[25]For some bibliography on the recent developments in artificial chiral materials see, for example, Bertolotti M, Sibilia C, Belardini A (2015) J Opt Soc America B32:1287

1.12 Gaussian Beams

A particular solution of Helmoltz equation is a kind of beams with axial symmetry called *Gaussian beams*. Gaussian beams are characterized by a Gaussian shape of the transversal profile of the beam. Considering a beam propagating in the z-direction, the electric field at any position z is given in the (x,y) plane as

$$\begin{aligned} E &= E_o(w_o/w)\exp\left\{-i(kz+\psi)-(x^2+y^2)(ik/2)\left[(1/R)-2i/kw^2\right]\right\} \\ &= E_o(w_o/w)\exp\left\{-\left[i(kz+\psi)-ik(x^2+y^2)/2R\right]\right\}\exp\left[-(x^2+y^2)/w^2\right] \end{aligned} \tag{1.119}$$

where E_o is the electric field which has a maximum at z=0, w_o is the minimum beam width at z = 0, called *beam waist*, connected to the beam width at a distance z by

$$w(z) = w_o\left[1+(z\lambda/w_o^2 n\pi)^2\right]^{1/2}. \tag{1.120}$$

The Gaussian beam is characterized by having a plane wave shape at z = 0 which gradually changes to a curved wave front of radius R as z increases.

At the point of intersection of the wave front with the z axis, the radius of curvature is

$$R(z) = z\left[1+(w_o^2 k/2z)^2\right]. \tag{1.121}$$

The spot size w(z) is a hyperbolic function around z = 0 with the minimum w_o at this position and an almost linear increase of beam radius at large distances (Fig. 1.13).

In fact from (1.120) we have

$$w(z) \sim 2z/kw_o \tag{1.122}$$

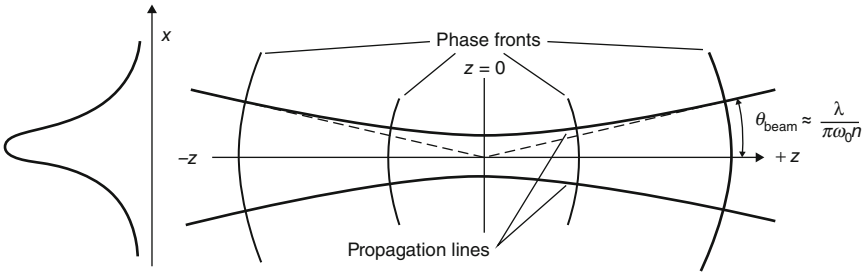


Fig. 1.13 A representation of a Gaussian beam. On the left the transverse profile, on the right the shape of the beam as a function of z in the plane xz

The angle of divergence as z becomes very large is

$$\theta_{beam} = w(z)/R = 2/kw_o \tag{1.123}$$

being R→z at large distances.

It is convenient to introduce a complex beam parameter q defined as

$$1/q = (1/R) - 2i/kw^2. \tag{1.124}$$

In terms of q one may write

$$\begin{aligned} E &= E_o(w_o/w_z)\exp\left[(r^2\pi/\lambda)(1/R - i\lambda/\pi w^2)\exp\{-i[kz - \varphi(z)]\}\right. \\ &= E_o(w_o/w_z)\exp\left[(r^2\pi/\lambda)q(z)\exp\{-i[kz - \varphi(z)]\}\right. \end{aligned} \tag{1.125}$$

The parameter q is useful to see how the gaussian beam is changed when traversing optical components, like mirrors or lens or other objects which can be described by a transport matrix of the kind

$$M = \begin{vmatrix} A & B \\ C & D \end{vmatrix}. \tag{1.126}$$

It can be shown that in this case the new value q′ of q is given by[26]

$$q' = (qA + B)/(qC + D). \tag{1.127}$$

The Gaussian beam is the typical beam emitted by a good laser.

Another useful parameter is the Rayleigh length defined as

$$z_r = (n\pi/\lambda_o)w_o^2. \tag{1.128}$$

It is a parameter that measures how far the beam is collimated. The smaller the beam waist, the shorter is the Rayleigh length.

At the Rayleigh length the beam radius is increased by a factor of $(2)^{½}$ compared to the waist value, and the intensity is reduced by a factor of 2. A plot of w(z) as a function of z is shown in Fig. 1.14. Thus up to a distance $z = z_r$ the spot size is approximately constant.

The phase ψ(z) in (1.119) is defined as

$$\psi = \tan^{-1}(z/z_R). \tag{1.129}$$

[26]See for example Yariv A (1975) Quantum electronics, 2nd edn. Wiley, New York

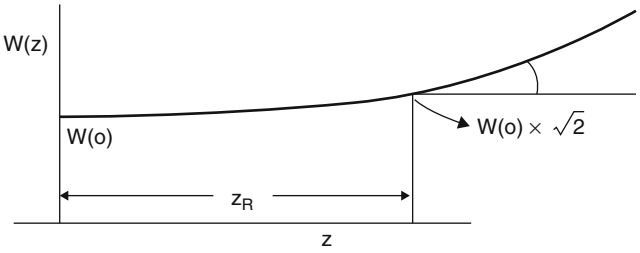


Fig. 1.14 Defining the Rayleigh length z_R

A plot of ψ as a function of z shows that the beam suffers a jump of 180° passing through the point $z = 0$ relative to the phase an ideal plane wave would have. It means that if the beam at $z \to -\infty$ was in phase with a reference plane wave, its phase will change gradually as the beam propagates through its focus so that the beam will be exactly out of phase with the reference wave at $z \to +\infty$. This behaviour is called *Gouy phase shift*.

Gaussian beams, like planar waves, are approximations. A true Gaussian beam, like a true planar wave, does not exist. The reason is that a true Gaussian beam should have a Gaussian profile along the axis and this profile extends to infinity so that a true Gaussian beam would have an infinite energy.

There exist also other kinds of Gaussian beams with a more complex field distribution in the plane xy.[27]

[27]See for example N Hodgson and H Weber, Optical Resonators, Fundamentals, Advanced Concepts and Applications, Springer 1997

Chapter 2
Evanescent Waves

2.1 Introduction

Evanescent waves or non-propagating inhomogeneous fields are an essential part of near-field optics. The most elementary near-field is the one associated with a single evanescent wave generated by total internal reflection at the surface of a dielectric material. It may be introduced considering the reflection and refraction of plane waves. Evanescent waves are also produced in the phenomena of diffraction.

Evanescent waves were investigated both experimentally[1] and theoretically[2] for long time, but have acquired great importance only in recent times.[3] They are at the base of all the material presented in this book.

2.2 Reflection and Refraction

We will briefly recall the principal facts about reflection and refraction of plane waves from a planar interface. This is the simplest way to introduce evanescent waves.

[1]Quincke G (1866) Ann Phys Chem. 127:1; Hall EE (1902) Phys Rev. 15:73; Selenyi P (1913) CR. 157:1408; Froehlich P (1921) Ann Phys. 65(4):577; Coon DD (1966) Am J Phys. 34:240; Drexhage KH (1970) Sci Am. 222:108; Nassenstein H (1968) Phys Lett. 28A:249; Bryngdahl O (1969) J Opt Soc America 59:1645; Smartt RN (1970) Appl Opt. 9:970

[2]Bouwkamp CJ (1959) Rep Progr Phys. 17:269; Wolf E (1959) Proc Phys Soc. (London) 74:269; Toraldo di Francia G (1960) Nuov Cim. 16:61; Sherman GC (1967) J Opt Soc America 57:1160; Sherman GC (1967) J Opt Soc. America 57:1490; Shewell JR, Wolf E (1968) J Opt Soc America 58:1596; Lalor E (1968) J Opt Soc America 58:1235; Walther A (1968) J Opt Soc America 58:1256; Walther A (1969) J Opt Soc America 59:1325; Asby R, Wolf E (1971) J Opt Soc America 61:52; Clemmow PC (1966) The plane wave spectrum representation of electromagnetic fields. Pergamon, New York

[3]de Fornel F (2001) Evanescent waves, from Newtonian optics to atomic optics. Springer

© Springer International Publishing AG 2017

M. Bertolotti et al., *Evanescent Waves in Optics*, Springer Series in Optical Sciences,
DOI 10.1007/978-3-319-61261-4_2

2.2.1 The Reflection and Refraction Cartesio-Snell Laws

The reflection of a plane electromagnetic wave at a planar interface between two media is completely characterized when solutions for two mutually perpendicular polarizations are known. The polarizations conventionally chosen are: one with its electric vector perpendicular to the plane of incidence (labelled s, from the German *senkrecht*, perpendicular), and the other with its electric vector parallel to the plane of incidence (labelled p).

Let us consider a plane wave incident on a plane surface between two homogeneous dielectric media of constants ε_1, μ_1 and ε_2, μ_2, respectively. The plane $z = 0$ is chosen to coincide with the separation plane between the two media. We call with θ the angle between the propagation direction of the incident wave and the axis z (we call it the *incidence angle*) and we chose the frame in such a way that the propagation vector lies in the plane xz that is called the *incidence plane*. In general two waves appear, both lying in the incidence plane: a **reflected wave** which travels in the same semi-space of the incident wave, characterized by a propagation vector that makes an angle θ' (called *reflection angle*) with the z-axis and a **refracted wave** that travels in the space under the plane, characterized by a propagation vector which forms an angle θ'' (*refraction angle*) with the z-axis (see Fig. 2.1).

The electric field of the three waves can be written as:

$$\mathbf{E} = \mathbf{E_o}\sin(\omega t - \mathbf{k}\cdot\mathbf{r}) \tag{2.1a}$$

for the incoming wave.

$$\mathbf{E}' = \mathbf{E_o'}\sin(\omega t - \mathbf{k}\cdot\mathbf{r}' + \delta') \tag{2.1b}$$

for the reflected wave, and

$$\mathbf{E}'' = \mathbf{E_o''}\sin(\omega t - \mathbf{k}''\cdot\mathbf{r} + \delta'') \tag{2.1c}$$

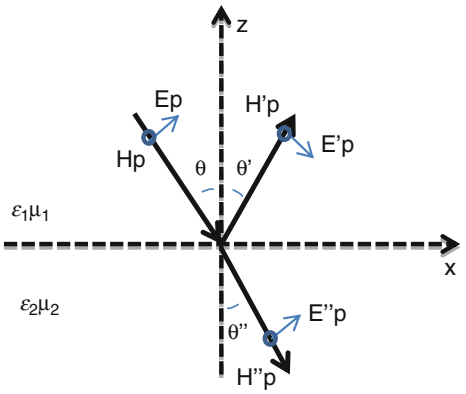


Fig. 2.1 Showing the geometry for reflection and diffraction for p-polarization

for the refracted wave. By applying the conservation law for the tangential component of the electric field at the boundary between the two media we derive the Cartesio-Snell's laws

$$\mathbf{k} \cdot \mathbf{r} = \mathbf{k} \cdot \mathbf{r}' = \mathbf{k}'' \cdot \mathbf{r}''. \tag{2.2}$$

Equation (2.2) shows that the reflected and refracted waves lie on the incidence plane and the phases δ',δ'' can be either 0 or π.

From the first equality of (2.2) it results

$$k \sin\theta = k \sin\theta' \tag{2.3}$$

which shows that $\theta = \theta'$, that is the **reflection law**.

The second equality in (2.2) gives

$$k \sin\theta = k'' \sin\theta'' \quad \text{or} \quad n_1 \sin\theta = n_2 \sin\theta'' \tag{2.4}$$

that is the **refraction law**.

Note that if we consider a photon, it may be either reflected or transmitted. In both cases the conservation laws of energy and momentum should be satisfied.

For reflection the conservation of energy is

$$\hbar\omega = \hbar\omega' \tag{2.5}$$

that shows that the frequency of incident and reflected beam must be the same, as we assumed from the beginning. For the conservation of momentum we remember that the momentum of a photon is $\hbar\omega/c = \hbar k$. The conservation of the x component of the propagation vector k is (2.3). The z component of the incoming wave k vector is $k\cos\theta$ and the z component of the reflected wave is $-k\cos\theta$. Therefore the conservation of momentum along the z-direction should be written

$$k\cos\theta = -k\cos\theta + p \tag{2.6}$$

where p is the contribution of the medium to the conservation and is

$$p = 2k\cos\theta. \tag{2.7}$$

This quantity is just the momentum exchanged with a perfectly reflecting surface by the photon assimilated to a molecule of a perfect gas colliding with a perfectly reflecting wall. In the case of molecules this momentum gives origin to the gas pressure. In the case of photons this originates the radiation pressure.

In the case of refraction the conservation of energy is again (2.5), with $\omega' = \omega''$, so that also the frequency of the refracted beam must be equal to that of the incoming beam. The conservation of the x component of the momentum is (2.4), while for the components of the k vectors in the z-direction we have $k\cos\theta$ and

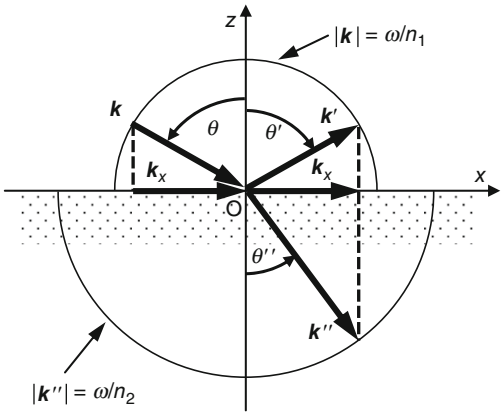


Fig. 2.2 Geometrical construction to find the reflected k′ and refracted k″ vectors, from the incident k.

k″cos θ″, for the incoming and refracted wave, respectively. Conservation of the momentum in the z-direction means therefore

$$k\cos\theta = k''\cos\theta'' + p''. \tag{2.8}$$

In this case p″ is the momentum exchanged with the virtual oscillators of the medium in the case of an elastic collision (because energy is conserved).

A geometric construction of the reflected and refracted ray may be made with reference to Fig. 2.2. Assume the incidence plane is the xz plane of the figure, the interface being the plane z = 0, with the incident medium in the positive semi-space. We draw the incident k vector forming the angle θ with the z-axis in the point O (x = 0, z = 0). We trace the semi-circle of radius $k = \omega/n_1$ and center in O. The perpendicular from the intercept of k with the semi-circle is k_x. If we write a segment equal to k_x from the other side and draw the normal from its apex we encounter the k semi-circle in a point which defines the reflected k′. To find the refracted k, that we label k″ we may draw from O a semi-circle of radius $k'' = \omega/n_2$. Drawing the perpendicular from the apex of k_x, it encounters this circle in a point which defines k″.

2.2.2 *Extension to Surfaces with a Phase Gradient*

More recently Capasso and collaborators[4] have extended Cartesio-Snell's laws to the case that along the boundary surface between the two media, a phase gradient exists, dΨ/dx. Such a situation can be obtained, for example, by artificially structuring the interface between the two media in order to introduce an abrupt phase shift in the light path, which is a function of the position along the interface.

[4]Yu N, Genevet P, Kats MA, Aieta F, Tetienne J-p, Capasso F, Gaburro Z (2011) Science 334:333

In this case we may still apply the conservation of momentum in the x direction, but we have to add the momentum component generated by the phase gradient, that is (dΨ/dx) to (2.4) that now reads

$$k\sin\theta + d\Psi/dx = k''\sin\theta'' \tag{2.9}$$

or

$$n_1\sin\theta = n_2\sin\theta'' - (\lambda o/2\pi)d\Psi/dx. \tag{2.10}$$

Equation (2.10) implies that the refracted beam can have an arbitrary direction, provided a suitable gradient of phase along the interface be introduced. Moreover, because of the nonzero phase gradient in (2.10), the two angles of incidence $\pm\theta$ lead to different values for the angle of refraction.

Similarly for reflection we have

$$\sin\theta = \sin\theta' - (\lambda o/2\pi n_1)d\Psi/dx. \tag{2.11}$$

An abrupt phase shift Ψ over the scale of the wavelength can be introduced in the optical path by suitably engineering the interface between two media, as shown in the paper by Capasso et al.

2.2.3 *Fresnel Coefficients*

The reflection and transmission coefficients (the so-called *Fresnel coefficients*) in the simple case of a constant boundary between two different media can be derived by applying the conservation of the tangential component of the electric E and magnetic H fields.

If we consider p polarization (Fig. 2.1), conservation of E means

$$E_p\cos\theta - E_p^{'}\cos\theta = E_p^{''}\cos\theta''. \tag{2.12}$$

In a similar way, considering that $B_y = E_x/v$, conservation of H, gives

$$B_y/\mu_1 + B_y^{'}/\mu_1 = B_y^{''}/\mu_2 \tag{2.13}$$

or

$$E_p/\mu_1 v_1 + E_p^{'}/\mu_1 v_1 = E_p^{''}/\mu_2 v_2. \tag{2.14}$$

Assuming $\mu_1 = \mu_2$ we may write (2.14) as

$$n_1(E_p + E_p^{'}) = n_2 E_p^{''}. \tag{2.15}$$

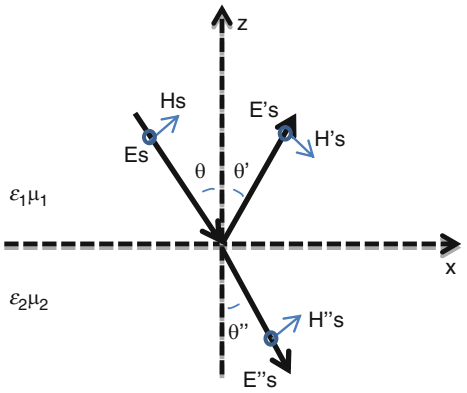


Fig. 2.3 The case of s-polarization in which the E vector is normal to the incidence plane

For waves polarized orthogonally with respect to the incidence plane we write (Fig. 2.3)

$$E_s + E_s^{'} = E_s^{''} \tag{2.16}$$

and

$$E_s \cos\theta/\mu_1 v_1 - E_s^{'} \cos\theta/\mu_1 v_1 = E_s^{''} \cos\theta''/\mu_2 v_2 \tag{2.17}$$

or (remember $\mu_1 = \mu_2$)

$$n_1 (E_s - E_s^{'}) \cos\theta = n_2 E_s^{''} \cos\theta''. \tag{2.18}$$

By simple algebra one eventually obtains

$$E_p^{'} = E_p tg(\theta - \theta'')/tg(\theta + \theta'') \tag{2.19}$$

and

$$E_p^{''} = E_p 2\sin\theta'' \cos\theta/\sin(\theta + \theta'')\cos(\theta - \theta''). \tag{2.20}$$

Similarly

$$E_s^{'} = -E_s \sin(\theta - \theta'')/\sin(\theta + \theta'') \tag{2.21}$$

and

$$E_s^{''} = E_s 2\sin\theta'' \cos\theta/\sin(\theta + \theta''). \tag{2.22}$$

Alternatively, by using the refractive index one may write with $n = n_2/n_1$

$$E_s' = E_s\left[\cos\theta - (n^2 - \sin^2\theta)^{1/2}\right] / \left[\cos\theta + (n^2 - \sin^2\theta)^{1/2}\right] \tag{2.23}$$

$$E_p' = E_p\left[n^2\cos\theta - (n^2 - \sin^2\theta)^{1/2}\right] / \left[n^2\cos\theta + (n^2 - \sin^2\theta)^{1/2}\right] \tag{2.24}$$

for the reflected waves, and

$$E_p'' = E_s 2\cos\theta / \left[\cos\theta + (n^2 - \sin^2\theta)^{1/2}\right] \tag{2.25}$$

$$E_p'' = E_p 2n\cos\theta / \left[n^2\cos\theta + (n^2 - \sin^2\theta)^{1/2}\right] \tag{2.26}$$

for the transmitted (refracted) waves.

We wish now discuss how the reflected and transmitted waves behave as a function of the angle of incidence. To do this we shall consider separately the case $n_1 < n_2$ and $n_1 > n_2$.

2.2.3.1 Reflection and Refraction for $n_1 < n_2$

Let us consider first the case $n_1 < n_2$. In this case we may define an *amplitude reflection coefficient* as

$$r = -E'/E \tag{2.27}$$

and an *intensity reflection coefficient* that is

$$R = |r|^2 = |E'|^2/|E|^2 = I'/I. \tag{2.28}$$

For p polarization we have

$$\begin{aligned} r_p &= -E_p'/E_p = -\mathrm{tg}(\theta - \theta'')/\mathrm{tg}(\theta + \theta'') \\ &= -\left[n^2\cos\theta - (n^2 - \sin^2\theta)^{1/2}\right] / \left[n^2\cos\theta + (n^2 - \sin^2\theta)^{1/2}\right] \end{aligned} \tag{2.29}$$

where $n = n_2/n_1$. The behaviour of r_p is shown in Fig. 2.4.

In the particular cases of $\theta = 0$ and $\theta = 90°$, r_p assumes the limit values

$$r_p(\theta = 0) = -(n-1)/(n+1) \quad \text{and} \quad r_p(\theta = 90°) = 1, \tag{2.30}$$

respectively passing through a zero value at some angle θ_p that is easily found putting

$$\mathrm{tg}(\theta_p + \theta'') = \infty \quad \text{or} \quad \theta'' = \pi/2 - \theta_p. \tag{2.31}$$

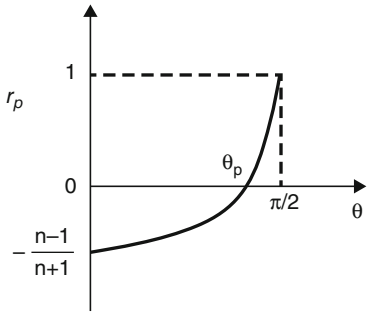


Fig. 2.4 The behaviour of r_p as a function of the incidence angle θ

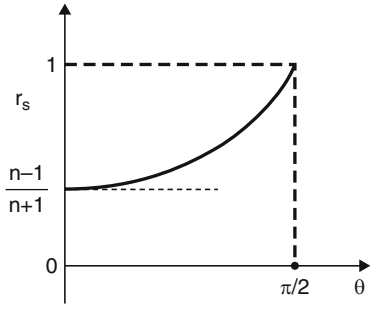


Fig. 2.5 The behaviour of r_s as a function of the incidence angle θ

from which it comes

$$\mathrm{tg}\theta_p = n. \tag{2.32}$$

The angle θ_p is called *Brewster angle* and under this incidence angle the reflected light is totally polarized in the s state (being the p amplitude zero).

For s polarization one has

$$\begin{aligned} r_s &= -E_s^{'}/E_s = \sin(\theta-\theta'')/\sin(\theta+\theta'') \\ &= -\left[\cos\theta - (n^2-\sin^2\theta)^{1/2}\right]/\left[\cos\theta + (n^2-\sin^2\theta)^{1/2}\right]. \end{aligned} \tag{2.33}$$

This reflection coefficient is shown in Fig 2.5.

We see that for $\theta = 0$ it is

$$r_s = (n-1)/(n+1) \tag{2.34}$$

and the coefficient steadily increases to reach the value 1 for $\theta = \pi/2$.

In a similar way we may define transmission coefficients that result as

$$t_p = E''/E = 2\sin\theta''\cos\theta/\sin(\theta+\theta'')\,\cos(\theta-\theta'') \tag{2.35}$$

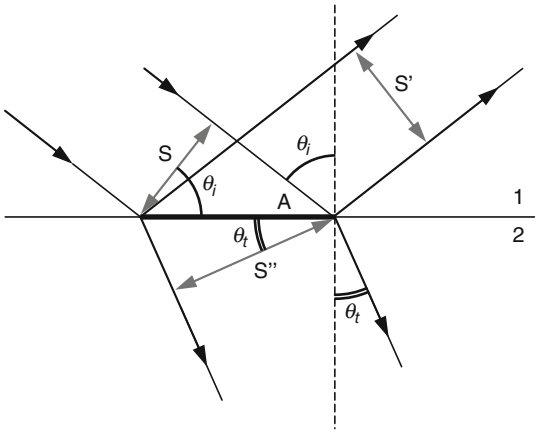


Fig. 2.6 showing the relation between the sections of the beams

and

$$t_s = E''/E = 2\sin\theta''\cos\theta/\sin(\theta+\theta''). \tag{2.36}$$

We want to discuss briefly the phase shift suffered in reflection. For $n_1 < n_2$ for s polarization there is a 180° phase shift for all angles, and for p polarization there is a 180° phase shift for angles below Brewster's angle and 0° for larger angles. At normal incidence $\theta = 0$ the reflection coefficients are <1. This can be understood considering that there will be destructive interference between the incident and reflected beams near the surface, where they overlap in space.

The power reflected and transmitted is

$$R = |E'|^2/|E|^2 = |r|^2 \tag{2.37a}$$

$$T = (|E''|^2/|E|^2)(n_2\cos\theta''/n_1\cos\theta) = |t|^2(n_2\cos\theta''/n_1\cos\theta). \tag{2.37b}$$

The reason for the term $(n_2\cos\theta''/n_1\cos\theta)$ in (2.37b) originates from the definition of intensity (eq. 1.16) and observing (Fig. 2.6) that the section of the beams on the boundary plane A are connected to the sections S, S′, S″ of the incident, reflected and transmitted beams by

$$S' = S = A\cos\theta \quad \text{and} \quad S''\cos\theta'' = A\cos\theta \tag{2.38}$$

It is easy to see that

$$R + T = 1 \tag{2.39}$$

for all polarizations.

In general for a wave polarized in an arbitrary direction, one may write the field as the vector sum of a part polarized in the p direction plus the part polarized in

the s direction. Because the reflection and transmission coefficients are different for the two polarizations in general the input polarization is not preserved in the reflected and transmitted waves. It results

$$R = I'/I = R_p(I_p'/I_p) + R_s(I_s'/I_s) \tag{2.40a}$$

$$T = (I''/I)(n_2\cos\theta''/n_1\cos\theta) = T_p(I_p''/I_p) + T_s(I_s''/I_s) \tag{2.40b}$$

where I are intensities.

A compact way to relate the amplitudes and polarizations of the incident, reflected and transmitted waves can be achieved using the Jones matrices

$$J_1 = \begin{vmatrix} E_p \\ E_s \end{vmatrix} \quad J_2 = \begin{vmatrix} E_p' \\ E_s' \end{vmatrix} \quad J_3 = \begin{vmatrix} E_p'' \\ E_s'' \end{vmatrix}. \tag{2.41}$$

They are related by the equations

$$J_2 = r\,J_1 \quad \text{and} \quad J_3 = tJ_1, \tag{2.42}$$

where r and t are 2 x 2 Jones matrices describing the transmission and reflection of the wave. The independence of the p and s polarizations implies that the Jones matrices r and t are diagonal

$$r = \begin{vmatrix} r_s & 0 \\ 0 & r_p \end{vmatrix} \quad \text{and} \quad t = \begin{vmatrix} t_s & 0 \\ 0 & t_p \end{vmatrix}. \tag{2.43}$$

We may now consider the most interesting case of $n_1 > n_2$ in the next subsection.

2.2.3.2 The Case $n_1 > n_2$

In the case that $n_1 > n_2$ one may put $n' = n_1/n_2 = 1/n$.

Defining an amplitude reflection coefficient as

$$r = E'/E \tag{2.44}$$

we find that

$$r_s = \sin(\theta'' - \theta)/\sin(\theta'' + \theta) = [n'\cos\theta - \cos\theta'']/[n'\cos\theta + \cos\theta] \tag{2.45}$$

$$r_p = \mathrm{tg}(\theta'' - \theta)/\mathrm{tg}(\theta'' + \theta) = [-n'\cos\theta'' + \cos\theta]/[n'\cos\theta'' + \cos\theta]. \tag{2.46}$$

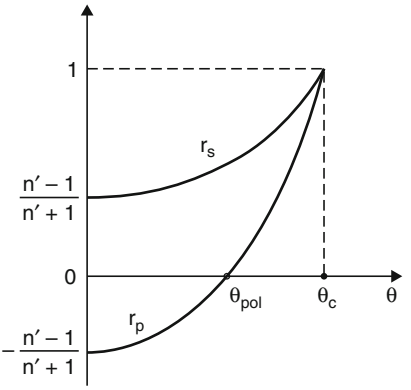


Fig. 2.7 Behaviour of the two reflection coefficients r_s and r_p as a function of the incidence angle θ

The behaviour of these two coefficients is shown in Fig. 2.7. Also in this case for p polarization there is an angle θ_{pol} at which the transmitted wave in this polarization has zero amplitude. This angle is given by

$$\mathrm{tg}\theta_{pol} = 1/n'. \tag{2.47}$$

But the most interesting point is that for both polarizations the maximum reflectivity $r = 1$ is reached at an angle θ_c that is less than $\pi/2$. This is a consequence of the refraction law (2.4) and of being $n_1 > n_2$. In fact in this case the refraction angle is larger than the incident angle and when

$$(n_1/n_2)\sin\theta_c = 1 \tag{2.48}$$

there are no real solutions for angles $\theta > \theta_c$. For angles larger than θ_c we have total reflection of the incident beam. The phase shift for s polarization in the region between 0 and θ_c is 0°. For p polarization the phase shift is zero for angles below Brewster angle and 180° for larger angles, up to the critical angle.

2.3 Evanescent Waves

We may assume that for $\theta > \theta_c$ the angle θ'', that at $\theta = \theta_c$ is $\theta'' = \pi/2$ becomes complex so that[5]

$$\sin\theta'' = \sin[(\pi/2) \pm i\beta] = \cos(\pm i\beta) = \cosh\beta > 1. \tag{2.49}$$

[5]Because $e^{i\alpha} = \cos\alpha + i\sin\alpha$ one has $\cos\alpha = (e^{i\alpha} + e^{-i\alpha})/2$ and if $\alpha = i\beta$, it is $\cos i\beta = (e^{-\beta} + e^{+\beta})/2 = \cosh\beta$

Using (2.49) one may write

$$r_s = \sin[(\pi/2) \pm i\beta - \theta]/\sin[(\pi/2) \pm i\beta + \theta] = \cos(\theta \mp i\beta)/\cos(\theta \pm i\beta) \tag{2.50}$$

whose modulus is 1 so that we may write

$$r_s = e^{i\gamma}. \tag{2.51}$$

Similarly it is

$$r_p = \mathrm{tg}[(\pi/2) - \theta \pm i\beta]/\mathrm{tg}[(\pi/2) + \theta \pm i\beta] = e^{i\delta}. \tag{2.52}$$

The results (2.51) and (2.52) show that in both polarizations when the incidence angle is larger than the critical angle, the modulus of the reflection coefficient is unity (total reflection). The reflected wave is therefore totally reflected and acquires a phase γ or δ different for the two polarizations given by

$$\mathrm{tg}(\gamma/2) = \sin\theta \sinh\beta/\cos\theta \cosh\beta = \left[(n')^2 \sin^2\theta - 1\right]^{1/2}/n'\cos\theta \tag{2.53}$$

$$\mathrm{tg}(\delta/2) = -n'\left[(n')^2 \sin^2\theta - 1\right]^{1/2}/\cos\theta \tag{2.54}$$

which is a function of the incidence angle θ.

One may ask what happens in the optically less dense medium for angles larger than the limit angle. Here they come the *evanescent waves*.

We saw that the incoming plane wave is reflected and refracted as plane wave. Let us write the field of the refracted wave considering the geometry of Fig. 2.8

$$E'' = A''\exp i(\mathbf{k}'' \cdot \mathbf{r} - \omega t) = A''\exp i(k''x \sin\theta'' + k''z\cos\theta'' - \omega t) \tag{2.55}$$

where

$$\cos\theta'' = i\left[(n')^2 \sin^2\theta - 1\right]^{1/2}. \tag{2.56}$$

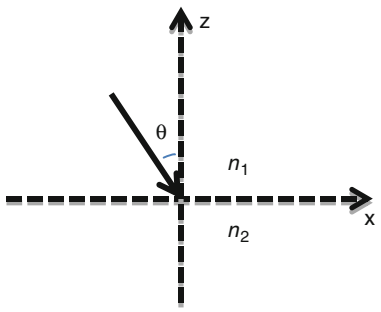


Fig. 2.8 The geometry for total reflection

Therefore

$$E'' = A''\exp\left\{-k''z\left[(n')^2\sin^2\theta - 1\right]^{½}\right\}\exp i(k''xn'\sin\theta - \omega t) = A''e^{-\alpha z}e^{i(kx-\omega t)} \tag{2.57}$$

where

$$\alpha = k''\left[(n')^2\sin^2\theta - 1\right]^{½} \tag{2.58}$$

and

$$k = n'k''\sin\theta. \tag{2.59}$$

A field, therefore, appears on the side of the less dense medium, and its intensity decays with distance z as $\exp(-\alpha z)$, with

$$\alpha = (2\pi n_2/\lambda_o)\left[(n_1/n_2)^2\sin^2\theta - 1\right]^{1/2} \tag{2.60}$$

where n_2 and n_1 are the refractive indices of less dense and the denser medium, respectively, θ is the angle of incidence, and λ_o is the vacuum wavelength.

Such a field is called *evanescent field* or *inhomogeneous field*. Equation (2.57) describes a field that propagates along the surface but decays exponentially into the less dense medium. Thus, a plane wave incident at an angle $\theta > \theta_c$ creates an *evanescent wave* in the medium with lower refractive index. Excitation of an evanescent wave in this way is referred as *total internal reflection* (TIR). For a glass/air interface at an angle of incidence $\theta_c = 45°$, the decay constant is $\alpha = 2.22/\lambda$. This means that already at a distance of about $\lambda/2$ from the interface, the time-averaged field is a factor of e smaller than at the interface. At a distance of about 2λ the field becomes negligible. The larger the angle of incidence the faster the decay will be.

The derivation of the amplitude A" can be made using the Fresnel formulae. For s polarization we may use (2.25) which in the case of an angle larger than the critical angle should be written

$$A'' = E_s'' = E_s 2\cos\theta / \left[\cos\theta + i(\sin^2\theta - n^2)^{½}\right] \tag{2.61}$$

with $n = n_2/n_1$, because the Fresnel coefficients for $\theta > \theta_c$ become complex numbers. Equation (2.61) can be written as

$$E_s'' = E_s 2\left[\cos\theta/(1-n^2)^{½}\right]\exp(-i\delta_s) \tag{2.62}$$

where the phase δ_s is

$$\mathrm{tg}\delta_s = (\sin^2\theta - n^2)^{1/2}/\cos\theta. \tag{2.63}$$

In a similar way for p polarization, from (2.26) one may derive for the two x and z components of the electric field

$$E_x'' = E_p 2\cos\theta(\sin^2\theta - n^2)^{1/2}/\left[n^4\cos^2\theta + \sin^2\theta - n^2)\right]\exp\left[-i(\delta_p + \pi/2)\right] \tag{2.64a}$$

$$E_z'' = E_p 2\cos\theta\sin^2\theta/\left[n^4\cos^2\theta + \sin^2\theta - n^2)\right]\exp(-i\delta_p) \tag{2.64b}$$

with

$$\mathrm{tg}\delta_p = (\sin^2\theta - n^2)^{1/2}/n^2\cos\theta. \tag{2.65}$$

The intensity of the p-polarized evanescent wave can be larger than that of the input beam, while for s polarized light no field enhancement is present.

The phase of the reflected and transmitted wave is shifted relative to the incident wave. This phase shift is the origin of the so-called *Goos–Hänchen shift*. Furthermore for p-polarized excitation, it results in elliptic polarization of the evanescent wave with a field rotating in the plane of incidence.

Evanescent waves never occur in a homogeneous medium but are inevitably connected to the interaction of light with inhomogeneities. The best known example where evanescent waves occur is the total internal reflection when, for example, a beam of light is incident from glass onto a glass-air interface at an angle greater than the critical angle for total internal reflection. This can be accomplished, for example, by directing a beam of light into a prism as shown in Fig. 2.9. The evanescent field may be seen approaching a transparent body to within less than $\lambda/2$ of the interface that supports the evanescent field. As shown in Fig. 2.9, this can be accomplished, for example, by using another prism or a sharp transparent fibre that converts the evanescent field at its tip into a guided mode propagating along the fibre. This is the basis for the optical scanning tunnelling microscopy as explained in Chap. 4.

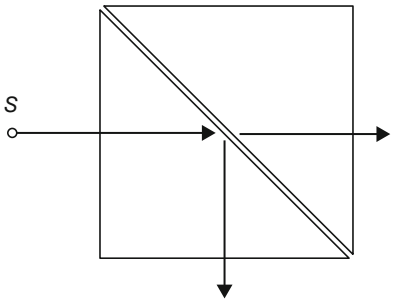


Fig. 2.9 Geometry to observe tunnelling of the evanescent wave

Newton[6] was the first to observe the evanescent waves. He describes an experiment where he contacted the facet of a prism illuminated under total reflection with another prism: "compressing two prisms hard together that their sides (which by chance were a very little convex) might somewhere touch one another, I found the place in which they touched to become absolutely transparent, as if they had there been one continued piece of glass."[7]

In the case of a surface with a phase gradient, there are two possible incidence angles $\pm\theta$ that lead to two different values for the angle of refraction. As a consequence there are two possible critical angles for total internal reflection, provided that $n_1 > n_2$

$$\sin\theta_c = \pm(n_2/n_1) - (\lambda_o/2\pi n_1)(d\Psi/dx) \tag{2.66}$$

and there is also a critical angle of incidence

$$(n_1/n_2)\sin\theta_c' = 1 - (\lambda_o/2\pi n_2)|d\Psi/dx| \tag{2.67}$$

above which the reflected beam becomes evanescent.[8]

2.4 Energy Transport by Evanescent Waves

For non-absorbing media and for incidence under an angle larger than the critical angle, all the power of the incident wave is reflected (*total internal reflection*, TIR). This means that there is no net energy transport into the medium with lower refractive index. This can be easily seen considering the z component of the time-averaged Poynting vector (Fig. 2.8)

$$<\mathbf{S}>_z = (½)\mathrm{Re}\left[(\mathbf{ExH}^*)_y - (\mathbf{ExH}^*)_x\right] \tag{2.68}$$

where the fields are evaluated in the medium into which the evanescent wave extends.

In the case of plane waves it is

$$\mathbf{H} = (\varepsilon_o\varepsilon_r/\mu_o\mu_r)^{½}[(\mathbf{k}/k)x\mathbf{E}]. \tag{2.69}$$

[6]Newton I (1726) Principia, vol. I. London; reprint of Cajori's English translation, Univ.Calif. Press 1962

[7]Newton I (1730) Opticks, 4th edn. London; Dover, New York, 1952

[8]Yu N, Genevet P, Kats MA, Aieta F, Tetienne J-p, Capasso F, Gaburro Z (2011) Science 334:333

Introducing the expressions for the transmitted field components of E and H into (2.68) one immediately see that $<S>_z$ vanishes and that there is no net energy transport in the direction normal to the interface.

On the other hand, when considering the energy transport along the interface it results

$$<S>_x = (½)(\varepsilon_2\mu_2/\varepsilon_1\mu_1)^{½}\sin\theta_1\left[|t_s|^2|E_{1s}|^2 + |t_p|^2|E_{1p}|^2\right]e^{-2\alpha x}. \tag{2.70}$$

Thus, an evanescent wave transports energy along the surface, in the direction of the transverse wave-vector. However, the absence of a net energy flow normal to the surface does not mean that there is no energy contained in an evanescent wave.

An evanescent field can interact, for example, with a molecule embedded in it and induce a transition that can be detected by the resulting fluorescence.

2.5 Tunnelling Effect

Evanescent waves can be converted into propagating radiation. This phenomenon is among the most important effects in near-field optics and finds a number of applications (for example in near-field optical microscopy). Considering the production of an evanescent wave in a prism of glass as shown in Fig. 2.9, if a second piece of glass is brought near the first piece, leaving an air gap between them, the internal reflection is "frustrated" and light can propagate into the second glass with a transmission coefficient near 1 if the width of the air gap is small compared with the wavelength. A demonstration of internal reflection has been given by Zhu.[9]

This situation is analogue to quantum mechanical tunnelling through a potential barrier.

Total reflection should not be considered simply a special case of partial reflection. The existence of a wave in the medium of lower refractive index has been proved in various experiments such as, for example, the tunnelling effect described above. In 1943 Goos and Hänchen[10] considered what happens at total reflection if the beam has a finite dimension, showing that the reflected wave suffers a spatial shift (*Goos–Hänchen shift*, Fig. 2.10) as it would happen if the incident wave penetrate into the medium of lower index and re-emerge into the medium of higher index. Evaluation of the Goos–Hänchen shift has been made by several authors under different approximations.[11]

[9] Zhu S et al (1986) Am J Phys. 54:601

[10] Goos F, Hänchen H (1947) Ann Physik 1:333

[11] See, for ex., Renard RH (1964) J Opt Soc America 34:1190; Lotsch HKV (1968) J Opt Soc America 58:551; Tamir T, Oliner AA (1969) J Opt Soc America 59:942; Horowitz BR, Tamir T (1971) J Opt Soc America 61:586

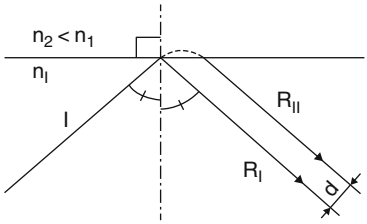


Fig. 2.10 Total reflection and the Goos–Hänchen shift. I is the incident beam. The reflected beam R_I shows the behaviour of a partially reflected beam; the reflected beam R_{II} shows the behaviour expected by Goos and Hänchen. The spacing d is the Goos–Hänchen shift

2.6 Reflection and Refraction in the Presence of Absorption

We wish now to discuss what happens if one or both dielectrics have absorption, that is have a dielectric permittivity and therefore a refractive index that has an imaginary part. To discuss this situation we describe the reflection and transmission phenomena using the wave equation.[12] Let us first start considering the case of s polarization and no absorption. We refer to Fig. 2.1 and write the wave equation for the y-component of the electric field

$$\partial^2 E_y/\partial x^2 + \partial^2 E_y/\partial z^2 + \varepsilon(z)(\omega^2/c^2)E_y = 0. \tag{2.71}$$

For planar interface, the dielectric function is only a function of z. The equation can therefore be solved separating the variables. We write

$$E_y(x, z) = X(x)Z(z). \tag{2.72}$$

Upon substitution and dividing for XZ we obtain

$$(1/X)\partial^2 X/\partial x^2 + (1/Z)\partial^2 Z/\partial z^2 + \varepsilon(z)\omega^2/c^2 = 0. \tag{2.73}$$

We may then put

$$(1/X)\partial^2 X/\partial x^2 = -K^2 \quad \text{and} \quad (1/Z)\partial^2 Z/\partial z^2 + \varepsilon(z)\omega^2/c^2 = K^2 \tag{2.74}$$

The solution of the first equation is simply

$$X = C\exp(iKx) \tag{2.75}$$

[12]In this discussion we follow the presentation given by Lekner J (1987) Theory of reflection, Martinus Nijhoff Publishers, Boston

where C is an arbitrary constant. The meaning of K is evident: it is the x component of the wave-vector k,

$$K = k_x = k'_x = n_1(2\pi/\lambda o)\sin\theta = n_1(2\pi/\lambda o)\sin\theta' = k''_x = n_2(2\pi/\lambda o)\sin\theta'' \quad (2.76)$$

Eq. (2.76) is Cartesio-Snell laws for reflection and diffraction.

The solution of the second equation, putting

$$q^2 = \varepsilon(z)\omega^2/c^2 - K^2 \quad (2.77)$$

is

$$Z = C'\exp(iqz) \quad (2.78)$$

with

$$q = q_1 = \varepsilon_1\omega^2/c^2 - K^2 \quad \text{in medium 1} \quad (2.79)$$

$$q = q_2 = \varepsilon_2\omega^2/c^2 - K^2 \quad \text{in medium 2.} \quad (2.80)$$

We immediately see that q_1, q_2 are the z component of the wave-vector in the two media, respectively

$$k_z = n_1\cos\theta \quad \text{and} \quad k'_z = n_2\cos\theta' \quad (2.81)$$

Note that the laws of reflection-refraction do not depend on the transition between the two media being sharp: they are valid for any arbitrary variation of ε (z) between the asymptotic values ε_1 and ε_2.

We may now find the reflection r and transmission t coefficients in the Fresnel formula. The E(z) part in medium 1 is given by

$$\exp(iq_1z) + r_s\exp(-iq_1z) \quad (2.82)$$

while it is $t_s\exp(iq_2z)$ in medium 2. In our case by imposing the continuity conditions for E(z) and its derivative dE(z)/dz at z = 0 one has

$$1 + r_{so} = t_{so} \quad \text{and} \quad iq_1(1 - r_{so}) = iq_2t_{so}. \quad (2.83)$$

From which it is easy to see that the Fresnell relations follow.

The case of p polarization is a bit more complex. In this case B =(0,By,0), and the wave equation is

$$\partial/\partial x(\varepsilon^{-1}\partial B_y/\partial x) + \partial/\partial z(\varepsilon^{-1}\partial B_y/\partial z) + \omega^2/c^2 B_y = 0. \quad (2.84)$$

Again we may separate the variables and write

$$B_y(x,z) = \exp i(Kx)B(z). \tag{2.85}$$

where K has the same meaning as in the s case. B(z) satisfies the ordinary differential equation

$$d/dz(\varepsilon^{-1}dB/dz) + (\omega^2/c^2 - K^2/\varepsilon)B = 0. \tag{2.86}$$

When ε is constant (outside the interfacial region) the p wave equation has the same form as the s wave equation, with the same wave-vector component q perpendicular to the interface. But within the interface there is an additional term proportional to the product of dε/dz and dB/dz. This term may be removed and (2.86) converted to the form of the s wave equation by operating a change of the z variable

$$Z = \varepsilon z. \tag{2.87}$$

Then the p wave equation reads

$$d^2B/dZ^2 + Q^2B = 0, \quad Q^2 = (1/\varepsilon)(\omega/c)^2 - K^2/\varepsilon^2. \tag{2.88}$$

Operating now as previously one finds[13]

$$-r_{po} = (Q_1 - Q_2)/(Q_1 + Q_2) \quad \text{and} \quad (\varepsilon_1/\varepsilon_2)^{1/2}\, t_{po} = 2Q_1/(Q_1 + Q_2) \tag{2.89}$$

where

$$Q_1 = q_1/\varepsilon_1 \quad \text{and} \quad Q_2 = q_2/\varepsilon_2 \tag{2.90}$$

From (2.89) the Fresnell formula are obtained. Also from (2.89) we see that the p wave shows zero reflection when $Q_1 = Q_2$, that is at the Brewster angle.

We now may extend the formalism to the case in which absorption is present. For the sake of simplicity let us assume that radiation is incident from a non-absorbing medium (real ε_1) onto an absorbing medium (complex $\varepsilon_2 = \varepsilon_r + i\varepsilon_i$).

For s wave the real and imaginary parts of q_2 are found from

$$q_2^2 = \varepsilon_2\omega^2/c^2 - K^2 = (\omega/c)^2(\varepsilon_r + i\varepsilon_i - \varepsilon_1 \sin^2\theta). \tag{2.91}$$

K is given by (2.76) and in this case the angle of refraction is complex and has a formal meaning only. Setting $q_2 = q_r + iq_i$, we have

$$q_r^2 - q_i^2 = (\omega/c)^2(\varepsilon_r - \varepsilon_1 \sin^2\theta) \quad ; \quad 2q_rq_i = (\omega/c)^2\varepsilon_i \tag{2.92}$$

[13]The reason for the term $(\varepsilon_1/\varepsilon_2)^{1/2}$ multiplying t_{po} is that the relation refers to the transmission of the electric field and not to the magnetic field

thus for $\varepsilon_i \neq 0$,

$$(cq_r/\omega)^2 = (1/2)\left\{\varepsilon_r - \varepsilon_1 \sin^2\theta + \left[(\varepsilon_r - \varepsilon_1 \sin^2\theta)^2 + \varepsilon_i^2\right]^{1/2}\right\} \tag{2.93}$$

$$cq_i/\omega = (\varepsilon_i/2)/(cq_r/\omega). \tag{2.94}$$

The waveform in the absorbing medium is

$$E_y(z, x, t) = \exp(-q_i z)\exp i(Kx + q_r z - \omega t). \tag{2.95}$$

For the sharp boundary between the two media that we are considering, the continuity of E and dE/dz at z = 0 gives

$$r_s = (q_1 - q_2)/(q_1 + q_2) = (q_1 - q_r - iq_i)/(q_1 + q_r + iq_i). \tag{2.96}$$

The intensity reflectivity is thus

$$R_s = \left[(q_1 - q_r)^2 + q_i^2\right]/\left[(q_1 + q_r)^2 + q_i^2\right]. \tag{2.97}$$

In a similar way we have

$$-r_p = (Q_1 - Q_2)/(Q_1 + Q_2) = (Q_1 - Q_r - iQ_i)/(q_1 + Q_r + iQ_i). \tag{2.98}$$

Where $Q_1 = q_1/\varepsilon_1$ and $Q_2 = q_2/\varepsilon_2$. Q_2 has a real and an imaginary part

$$Q_r = (\varepsilon_r q_r + \varepsilon_i q_i)/(\varepsilon_r^2 + \varepsilon_i^2) \quad \text{and} \quad Q_i = (\varepsilon_r q_i - \varepsilon_i q_r)/(\varepsilon_r^2 + \varepsilon_i^2). \tag{2.99}$$

The p intensity reflectivity is

$$R_p = \left[(Q_1 - Q_r)^2 + Q_i^2\right]/\left[(Q_1 + Q_r)^2 + Q_i^2\right]. \tag{2.100}$$

In the absence of absorption, the p wave reflectivity is zero for $Q_1 = Q_2$ (at the Brewster angle). In case of absorption it is not so. Thus perfect absorption at a single sharp boundary is not possible.

An important and dramatic effect due to absorption is that of *attenuated total reflection*. An absorbing layer (typically a metallic film) deposited between two dielectrics can turn a total reflection configuration into one whose p reflectance is small at resonance, and can be zero for proper choice of thickness of metal film and angle of incidence. This phenomenon is an interference attenuation effect, associated with the resonant excitation of electromagnetic surface waves at a metal-dielectric interface (Chap. 5).

2.7 Reflection and Refraction with Materials with Negative Refractive Index

Some years ago, Veselago[14] draw attention to the fact that a negative electric permittivity ε and a negative magnetic permeability μ implies a negative refractive index. At first thought, it would seem that having $\varepsilon_r < 0$ and $\mu_r < 0$ should not affect the refractive index $n = (\varepsilon_r \mu_r)^{1/2}$. However, in real life, ε_r, μ_r and n are complex and we can write[15]

$$\varepsilon = r_\varepsilon e^{i\theta} \quad \text{and} \quad \mu = r_\mu e^{i\varphi} \tag{2.101}$$

and

$$n = (\varepsilon_r \mu_r)^{1/2} = (r_\varepsilon r_\mu)^{1/2} e^{i(\theta+\varphi)/2}. \tag{2.102}$$

The requirement that the imaginary part of n be positive for a passive medium implies that

$$0 \le \frac{1}{2}(\theta + \phi) < \pi. \tag{2.103}$$

If the real parts of ε and μ are both negative, i.e. $\cos\theta < 0$ and $\cos\phi < 0$, then

$$\frac{\pi}{2} < \frac{1}{2}(\theta + \phi) < \frac{3\pi}{2}. \tag{2.104}$$

To satisfy both (2.103) and (2.104), we must have

$$\frac{\pi}{2} < \frac{1}{2}(\theta + \phi) < \pi \tag{2.105}$$

and, therefore, a negative (real) refractive index:

$$n_R \equiv \mathrm{Re}[n] = \sqrt{r_\varepsilon r_\mu}\cos\frac{1}{2}(\theta + \phi) < 0. \tag{2.106}$$

Thus, the requirement that n has a positive imaginary part leads to the conclusion that, if ε and μ have negative real parts, the real part of n must also be negative.

An immediate consequence of a negative refractive index can be seen from Cartesio-Snell's law. Consider a plane wave incident from a medium with $n_1 > 0$

[14] Veselago VG (1967) Sov Phys Solid State 8:2854; Veselago VG (1968) Sov Phys Uspeki 10:509

[15] An excellent text where the properties of negative index are synthesized is Milonni PW (2005) Fast light, slow light and left-handed light. IoP, Bristol; A good reference for metamaterials is Solymar L, Shamonina E (2009) Waves in metamaterials. Oxford Univ. Press, Oxford

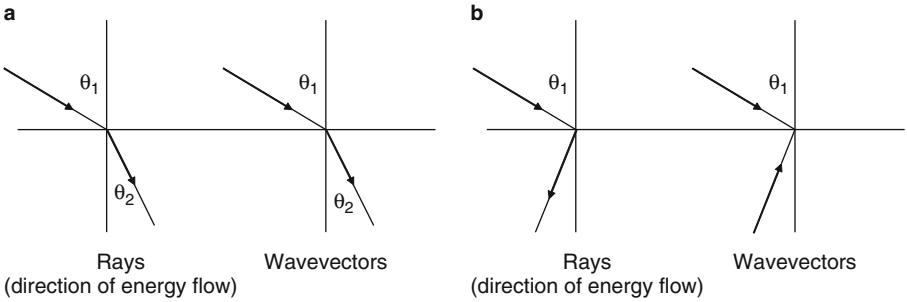


Fig. 2.11 The rays and vectors for refraction from vacuum to a material with (a) $n > 0$ and (b) $n < 0$. In the latter case the direction of the wave-vector and, therefore the phase velocity is opposite to the direction of energy flow

onto a negative-index medium with $n_2 < 0$ (NIM). If θ and θ'' are the angles of incidence and transmission, respectively, then according to Cartesio-Snell's law

$$n_1 \sin\theta = n_2 \sin\theta''. \tag{2.107}$$

$n_2 < 0$ implies that $\sin\theta'' < 0$, that is the transmitted rays make a *negative* angle with respect to the normal to the interface. In other words, the refracted rays are bent to the same side of the normal as the incident rays – the "wrong" way compared with the usual case of positive-index media. This is shown in Fig. 2.11.

Consider now a plane wave propagating in an NIM. Write

$$\boldsymbol{E} = \boldsymbol{E}_0 e^{i(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} \tag{2.108}$$

and, likewise, for D, B and H. From the Maxwell curl equations (1.1c) and (1.1d) with $j = 0$ and the constitutive relations (1.4a) and (1.4b) for an isotropic medium, we obtain

$$\boldsymbol{k} \times \boldsymbol{E}_0 = \omega\mu(\omega)\boldsymbol{H}_0 \tag{2.109}$$

and

$$\boldsymbol{k} \times \boldsymbol{H}_0 = -\omega\varepsilon(\omega)\boldsymbol{E}_0. \tag{2.110}$$

Thus if $\varepsilon(\omega) > 0$ and $\mu(\omega) > 0$, the vectors E, H and k form a right-handed triad and the Poynting vector $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ points in the direction of the wavevector $\mathbf{k}$. But if $\varepsilon(\omega) < 0$ and $\mu(\omega) < 0$, the vectors $\mathbf{E}$, $\mathbf{H}$ and $\mathbf{k}$ form a *left*-handed triad and the Poynting vector points in a direction *opposite* to the direction of the wavevector $\mathbf{k}$. For this reason, electromagnetic waves in a NIM may be called *left-handed*. In a NIM the group velocity (and the Poynting vector) and the phase velocity point in opposite directions.

Veselago noted also that frequency dispersion and absorption are necessary for $\varepsilon < 0$ and $\mu < 0$, since other ways the total energy would be negative.

It can also be shown that – away from any absorption resonances – the group index $n_g = n + \omega dn/d\omega$ is positive in a NIM.

In ordinary materials $\varepsilon, \mu < 0$ can occur, but the frequencies at which $\varepsilon < 0$ are generally very different from those at which $\mu < 0$. There are no known naturally occurring materials for which simultaneously $\varepsilon, \mu < 0$. However, simultaneously negative ε, μ are realized in artificial *metamaterials*. Metamaterials are so named because they are materials structured at a nanoscale able to have properties not attainable with ordinary materials.

Other interesting properties were described by Veselago: The Doppler effect is reversed in a negative index material and a spontaneously emitting atom will experience a recoil momentum in the same direction as the Poynting vector of the emitted light. Similarly, if light in a negative index material is incident on a reflecting interface, it will impart to the reflector a linear momentum in the direction opposite to the Poynting vector. Such strange consequences are perfectly consistent with the conservation of energy and linear momentum.

We are here interested to see how Fresnel formulas behave in the presence of a negative index material.

Consider the case in which light is incident from a positive-index medium onto a negative-index medium such that $n_2 < 0$. It is easy to show that the Fresnel equations for positive index media are applicable to negative as well as positive index media for $\theta < \theta_c$, for any polarization of the field, provided we replace all n, ε and μ in the formulas by their absolute values.

When the incident angle is larger than the critical angle, the square root in (2.56) defining the imaginary cosinus of the refraction angle was chosen positive for $n > 0$ because the transmitted field (2.57) must not diverge as z goes to infinity. For $n < 0$ the minus sign in (2.56) should be chosen for the same reason. The consequence is that the amplitude reflection coefficient r has now a phase angle of opposite sign with respect of the case with $n > 0$.

An important result is obtained when $n_2 = -n_1$. Fresnel formulas show that in this case there is no reflected light, no matter what the angle of incidence or the field polarization.

2.8 X-ray Evanescent Waves

For X rays the determination of the refractive index can be made by using the same formalism we have used in Chap. 1, writing

$$\begin{aligned} n(\omega) = 1 + (Ne^2/\varepsilon_o m)\Sigma f_i(\omega_{io}^2 - \omega^2)/\left[(\omega_{io}^2 - \omega^2)^2 + (\gamma_i\omega)^2\right] \\ + i(Ne^2/\varepsilon_o m)\Sigma f_i\gamma\omega/\left[(\omega_{io}^2 - \omega^2)^2 + (\gamma_i\omega)^2\right]. \end{aligned} \tag{2.111}$$

Considering that the frequency of X rays is very far from the resonant frequencies of the electrons this expression can be written

$$n(\omega) = 1 - \delta(\omega) + i\beta(\omega) \tag{2.112}$$

where

$$\delta(\omega) = (Ne^2/\varepsilon_o m)\Sigma f_i(\omega^2 - \omega_{io}^2)/\left[(\omega^2 - \omega_{io}^2)^2 + (\gamma_i\omega)^2\right] \tag{2.113}$$

$$\beta(\omega) = i(Ne^2/\varepsilon_o m)\Sigma f_i\gamma\omega/\left[(\omega^2 - \omega_{io}^2)^2 + (\gamma_i\omega)^2\right]. \tag{2.114}$$

For X rays $\beta << \delta$ and $\delta \sim 10^{-5}$, so that we may assume $\beta = 0$ which allows a very simple expression for n

$$n = 1 - \delta. \tag{2.115}$$

For an electromagnetic wave whose frequency ω exceeds that associated with the proper frequency of the majority of electrons in the medium ω_o (see chapter I), $n < 1$ (a phase velocity greater than c is obtained). The reflection then occurs externally with an evanescent field penetrating into the medium which now appears to be optically less dense than the vacuum. This is the case that applies to hard X rays on all solid targets and the possibility of total external reflection of such radiation has been known for some time.

In this case the vacuum has a refractive index $n_1 = 1$ while the medium has $n_2 = 1 - \delta + i\beta$. The critical angle for total refection from (2.39) results

$$\sin\theta_c = (1 - \delta). \tag{2.116}$$

The angle θ_c is very near to $\pi/2$ so that it is convenient to consider the grazing angle $\alpha = (\pi/2) - \theta$. Substituting in (2.116), we have

$$\alpha_c = (2\delta)^{1/2} \tag{2.117}$$

δ may be expressed as

$$\delta = ZNe^2\lambda^2/4\pi^2\varepsilon_o mc^2 \tag{2.118}$$

where Z is the atomic number, so that

$$\alpha_c = (ZN/\varepsilon_o m)^{1/2}(e\lambda/2\pi c). \tag{2.119}$$

Total reflection occurs for angles $\alpha < \alpha_c$. An estimate is α_c (rad) = $3.10^{-3}\lambda$ (Å). With $\lambda = 0.73$ Å it results α_c of the order of milliradiants.

The possibility of extending the utility of total reflection to study thin epitaxial films on crystals was realized[16] using total reflection conditions to select elastic Bragg scattering from surface only. A theoretical analysis was also made.[17]

A series of experiments designed to probe the x-ray evanescent wave properties was made by Becker et al.[18]

Diffraction produced by the evanescent X-ray field under the surface of a sample is used to investigate the atomic structure of surfaces and interfaces in a depth of a few atomic layers. The information obtained by evanescent diffraction is not merely surface-sensitive, but is unique compared with a conventional diffraction method because in evanescent diffraction, the scanning vector **k** is almost parallel to the surface and diffraction intensities are measured by scanning a detector in a plane parallel to the surface (in plane diffraction). Thus evanescent diffraction gives information on the two-dimensional periodic structure in the near-surface region and its change along the depth direction.

2.9 Reflection and Refraction of Plane Waves at a Boundary Between an Isotropic and a Birefringent Medium

In this case it is necessary to remember that in the birefringent medium in general two propagation directions are allowed with different velocity. The situation may be schematized as shown in Fig. 2.12.

The refraction law is the same as (2.4). However there are two big differences in the case of an anisotropic medium. The first is that in the second medium we do not know n and θ″. In fact n depends on the direction of k and in general for

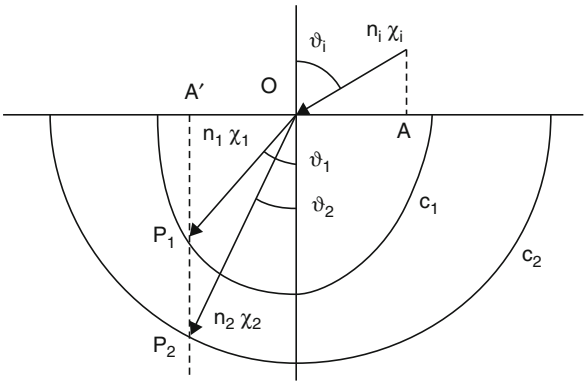


Fig. 2.12 Construction of the two refracted beams in a birefringent crystal

[16]Marra WC, Eisenberger P, Cho AY (1979) J Appl Phys. 50:6927

[17]Vineyard GH (1982) Phys Rev B26:4146

[18]Becker RS, Golovchenko JA, Patel JR (1983) Phys Rev Lett. 50:153

any direction of k two values of n exist. Equation (2.4) alone is therefore not enough to find the refracted waves. The second difference is that the Poynting vector may be non-collinear with k. Equation (2.4) refers to the state of the k vectors and therefore does not specify the direction of the flow of energy which may be not in the incidence plane.

A geometric construction which allows to solve the problem is shown in Fig. 2.12. In the incidence plane we draw the incident vector here indicated as $n_i\boldsymbol{\chi_i}$. The projection of this vector on the interface, OA is $n_i\sin\theta_i$. In the semi-space of the anisotropic medium we draw the two curved lines obtained considering for each direction the two k values that a plane wave propagating in that direction would encounter. In the figure these two curves are labelled c_1 and c_2. We consider now the point A′ symmetric of A with respect to O and draw the perpendicular to A′ to the interface between the two media. This line encounters the curves c_1 and c_2 into two points P_1 and P_2, respectively. The segment OP_1 and OP_2 give the directions of the two waves. Note that the projections of OP_1 and OP_2 over the interface are $n_1\sin\theta_1$ and $n_2\sin\theta_2$ both equal to $n_i\sin\theta_i$.

2.10 The Plane Wave Decomposition of a Field

Any electromagnetic field can be represented as a superposition of plane waves of different directions. Such an angular spectrum representation of electromagnetic field[19] consists in representing a given field as a superposition of plane waves of different propagation vectors. This representation yields a decomposition into homogeneous plane waves and inhomogeneous evanescent waves which appear as a natural adjunct to the spectrum of homogeneous plane waves.

To understand this point, suppose we expand the electric field in the half space $z > 0$ in terms of spatial and temporal frequency components

$$E(x,y,z,t) = \int_0^\infty d\omega \int_\infty^\infty dk_x \int_\infty^\infty dk_y A(k_x,k_y,\omega,z)e^{-i(\omega t - k_x x - k_y y)} \tag{2.120}$$

where, as usual, the real part of the right-hand side is implicit. If any sources of radiation are present only in the half-space $z < 0$, we have

$$\nabla^2 E - \frac{1}{c^2}\frac{\partial^2 E}{\partial t^2} = 0 \tag{2.121}$$

[19]Bouwkamp CJ (1954) Rept Progr Phys. 17:39; Wolf E (1959) Proc Phys Soc. (London) 74:269; Clemmow PC (1966) The plane wave spectrum representation of electromagnetic fields. Pergamon Press, New York; Sherman GC (1967), J Opt Soc America 57:1160; Sherman GC (1967), J Opt Soc America 57:1490; Sherwall JR, Wolf E (1968) J Opt Soc America 58:1596; Lalor E (1968) J Opt Soc America 58:1235; Goodman JW (1996) Introduction to Fourier Optics. McGraw, New York

for z > 0 and, therefore,

$$\frac{\partial^2 A}{\partial z^2} + \left(\frac{\omega^2}{c^2} - k_x^2 - k_y^2\right) A(k_x, k_y, \omega, z) = 0 \tag{2.122}$$

which has solutions in the form of plane waves

$$A(k_x, k_y, \omega, z) \propto e^{\pm i k_z z} \tag{2.123}$$

where the propagation constant is

$$k_z = \sqrt{\frac{\omega^2}{c^2} - k_x^2 - k_y^2}. \tag{2.124}$$

k_z can be real or imaginary, depending on whether $k_x^2 + k_y^2 < \omega^2/c^2$ or $k_x^2 + k_y^2 > \omega^2/c^2$. In the latter cased we have

$$k_z = i|k_z|. \tag{2.125}$$

Then

$$\begin{aligned} E(x, y, z, t) = & \int_0^\infty d\omega \iint_{k_x^2 + k_y^2 < \omega^2/c^2} dk_x dk_y A_{\text{ord}}(k_x, k_y, \omega) \\ & \times e^{i(k_x x + k_y y + k_z z - \omega t)} \\ & + \int_0^\infty d\omega \iint_{k_x^2 + k_y^2 > \omega^2/c^2} dk_x dk_y A_{\text{ev}}(k_x, k_y, \omega) \\ & \times e^{i(k_x x + k_y y - \omega t)} e^{-|k_z| z}. \end{aligned} \tag{2.126}$$

The first term on the right-hand side is a superposition of ordinary (homogeneous) plane waves propagating in the positive z direction. The second term defines a superposition of evanescent wave components that decay exponentially with z.

It is clear from the derivation of (2.126) that to describe the field E they are required for completeness. They are present whenever the field in some plane (z = 0) has spatial frequency components k_x, k_y such that $k_x^2 + k_y^2 > \omega^2/c^2$, i.e. whenever the field in this plane varies on a scale comparable to the wavelength or smaller.

The expansion (2.126) shows that the evanescent waves do not undergo any variation of phase in the z direction. This is related to the fact that evanescent waves, and they quantum mechanical analogue of tunnelling into a classically forbidden region, can exhibit "superluminal" behaviour. The only phase variation of the evanescent waves is along their propagation direction, parallel to the interface.

We remember that an important characteristic of evanescent waves is that they do not transport energy in the direction in which they decay. Because of their exponential decay with distance, evanescent waves are said to belong to the near

field of the object, as opposed to the far field associated with the homogeneous plane-wave components. Near-field optics (e.g. a scanning tunnelling microscope) involves distances close enough to an object that the evanescent waves are captured giving information about the object that is not available in the far field.

2.11 The Classical Limit of Resolution Explained

The light illuminating an object is scattered in any direction and this scattered light contains all the information on the object features. In the scattering process all the features that have a scale smaller than the wavelength of the incoming light originates evanescent waves that do not propagate freely in space but decay exponentially with distance so that if the object is observed at a distance of a few wavelengths or more these waves have disappeared and the features that they represented are irremediably lost. This is the reason why ordinary optical microscopes have a resolving power, that is the ability to distinguish as separate two near points, that is about equal to the wavelength used for the illumination. The near field optics (for example the scanning microscope) implies to be at a very near distance from the object so to capture the evanescent waves which give information on the object that are not available in the far field.

Considering the angular expansion of the field, we may say that large values of k_x and k_y in the field irradiated or reflected by a bi-dimensional object correspond to small spatial features, so that for good image resolution it is necessary to access high spatial frequencies.

Equation (2.124) suggest that the maximum possible resolution Δ is achieved when

$$k_x^2 + k_y^2 = k_{\max}^2 = k^2 \tag{2.127}$$

so that

$$\Delta = \frac{2\pi}{k_{\max}} = \frac{2\pi c}{\omega} = \lambda. \tag{2.128}$$

Implicit in this argument is the assumption that $k_{max} = k$, i.e. that $k_z^2 \geq 0$. In other words, the limit (2.128) assumes that the accessible field from the object has no evanescent components. Since the evanescent components of the field decay with distance z from the object as

$$e^{-|k_z|z} = \exp\left(-\sqrt{k_x^2 + k_y^2 - \omega^2/c^2}\right)z \tag{2.129}$$

the largest spatial frequency components decay most rapidly with z. In particular, spatial frequencies corresponding to spatial features on a scale of λ on the object do not appear in the field at distances greater than about λ from the object. For this reason,

the maximum resolution possible with a conventional lens is given by (2.128), regardless of how good the glass is, or how large the aperture is. Features on the scale of the wavelength or smaller cannot be resolved by an ordinary lens because an ordinary lens does not capture the evanescent components of the object field.

Pendry[20] has demonstrated that, using a planar lens with negative index (in particular it should be $\varepsilon_r = \mu_r = -1$), the evanescent waves grow with the distance from the lens and are transmitted. So that all components of spatial frequency may, in principle, contribute to the image.

The limit $\Delta \sim \lambda$ is not therefore a physical condition but rather an instrumental limit.

Coming back to the problem of resolution, we may also look at resolution from another point of view, considering Heisenberg's uncertainty principle

$$\Delta q \Delta p \geq h. \tag{2.130}$$

Remembering the relations

$$p = h/\lambda \quad \text{and} \quad k = \omega/c = 2\pi/\lambda \tag{2.131}$$

we have

$$\Delta q \Delta p = \Delta q \hbar \Delta k \geq h \tag{2.132}$$

or

$$\Delta q \Delta k \geq 2\pi \tag{2.133}$$

In one dimension this is

$$\Delta x \Delta k_x \geq 2\pi \tag{2.134}$$

or

$$\Delta x \geq \lambda \tag{2.135}$$

that is the previously found criterion.

Also in this case one may reduce Δx to $\Delta x < \lambda$ by squeezing the field just increasing the spread Δk_x.

The resolution limit was discussed by Abbe[21] and Lord Rutherford.[22] They did not pay attention to what happens in the near field being interested to the

[20] Pendry JB (2000) Phys Rev Lett. 85:3966

[21] Abbe E (1873) Archiv f Mikroskop Anat. 9:413

[22] Lord Rayleigh (1896) Phil Mag. 5(42):167

resolution limit that could be attained with optical instruments like telescopes or microscopes. Lord Rayleigh considering the diffraction pattern given by two near point sources proposed a limit $\Delta \sim \lambda$. Abbe, considering the resolution of a microscope objective derived $\Delta \sim \lambda/\sin\theta$, where θ is the objective aperture. These two formula are what to day is called the *classical resolution limit*.

2.12 Reflection and Refraction of Gaussian Beams

Let us consider a Gaussian beam incidents over an interface between two homogeneous media in the same geometry used for plane waves. To obtain solutions for the reflected and refracted fields, one may use the previously described technique of expanding the incident field in a Fourier integral, and consider every Fourier component as an independent plane wave. Every such wave generates one reflected plane wave in the first medium and one refracted plane wave in the second medium, whose parameters are determined by the well known laws of reflection and refraction for plane waves.[23]

The result is that the reflected and refracted beams maintain the Gaussian shape. However the reflected beam suffers a small displacement even for angles smaller than the critical angle. An approximate treatment which does not take this effect into account can be made rotating the coordinate frame[24] in which the incident, reflected and transmitted beams are represented so to made the beams impinge with the desired angles.

If θ is the incidence angle, the transformations to describe the incident beam are

$$x_i = x\cos\theta - z\sin\theta, \quad y_i = y, \quad z_i = x\sin\theta + z\cos\theta \tag{2.136}$$

and

$$r_i^2 = x_i^2/w_{ix}^2 + y_i^2/w_{iy}^2 = (x\cos\theta - z\sin\theta)^2/w_{ix}^2 + y^2/w_{iy}^2 \tag{2.137}$$

where we assume that the axis of the incident Gaussian beam passes through the origin and admit an elliptic beam shape with two radii w_{ix} and w_{iy}.

Similarly for the reflected beam it is

$$x_r = -x\cos\theta - z\sin\theta, \quad y_r = y, \quad z_r = -x\sin\theta - z\cos\theta \tag{2.138}$$

and

$$r_r^2 = x_r^2/w_{rx}^2 + y_r^2/w_{ry}^2 = (x\cos\theta - z\sin\theta)^2/w_{rx}^2 + y^2/w_{ry}^2 \tag{2.139}$$

[23] Horowitz BR, Tamir T (1971) J Opt Soc America 61:586

[24] McDonald KT, http://puhep1.princeton.edu/mcdonald/exemples/

and for the transmitted beam it is

$$x_t = x\cos\theta'' - z\sin\theta'', \quad y_t = y, \quad z_t = x\sin\theta'' + z\cos\theta'' \tag{2.140}$$

and

$$r_t^2 = x_t^2/w_{tx}^2 + y_t^2/w_{ty}^2 = (x\cos\theta'' - z\sin\theta'')^2/w_{tx}^2 + y^2/w_{ty}^2. \tag{2.141}$$

Applying the continuity conditions for E and H it results

$$w_{ix} = w_{rx} = (\cos\theta/\cos\theta'')w_{tx}, \quad w_{iy} = w_{ry} = w_{ty}. \tag{2.142}$$

Therefore, the reflected beam maintains the same shape as the incoming one, while the transmitted beam changes its shape.

In the case of going from a denser to a rarer medium, the abrupt transition between field propagation and exponential decay at the critical angle does not take place for the refracted Gaussian beam. In fact, in the spectrum of the Gaussian beam, there are plane waves with various directions of propagation, which collect around average directions of propagation of a beam as a whole. Therefore, when the incident angle of a beam is very close to the critical value, for one part of its plane-wave spectrum the incident angle is smaller than that, but for another part of spectrum it is greater. Summation over all contributions of those and others in the resulting field of the refracted beam should produce a continuous transition from beam propagation to amplitude decay.

The problem of reflection, refraction and scattering of Gaussian beams has been treated by several authors with different formalisms.[25]

2.13 Evanescent Waves in Diffraction

Evanescent waves are always produced in diffraction phenomena. This is implicit in the field expansion (2.126). We will here discuss a very simple example. Consider the case of a one dimensional amplitude sinusoidal grating that may be obtained making the transmission τ[26] of a thin glass plate to vary along the x direction according to the law

$$\tau = a + b\cos(2\pi x/L) \tag{2.143}$$

[25]Deschamps GA (1971) Electr Lett. **7**:684; Ra JW, Bertoni HL, Felsen LB (1973) SIAM J Appl Math. 24:396

[26]We may define the amplitude transmittance function as the ratio of the transmitted field amplitude to the incident field amplitude.

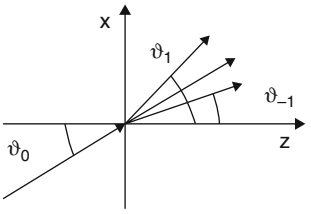


Fig. 2.13 Scattering of a plane wave from a sinusoidal amplitude grating lying along the x direction

where a and b are two suitable constants limited to satisfy the following conditions $a + b \leq 1$ and $a \geq b \geq 0$, and L is the spatial period. If a plane wave comes on the grating making an angle θ_o with the normal to the grating that we take the z direction (Fig. 2.13), we may write for the electric field on the $z = 0$ plane, where is laying the grating, dropping the time dependence

$$E = Ae^{i(kx\sin\theta)}. \tag{2.144}$$

Immediately behind the grating, the field can well be approximated by

$$\begin{aligned} E_{out} &= A\tau e^{i(kx\sin\theta o)} \\ &= A[a + b\cos(2\pi x/L)]e^{i(kx\sin\theta o)} \\ &= Aae^{i(kx\sin\theta o)} + (Ab/2)e^{i(k\sin\theta o + 2\pi/L)x} + (Ab/2)e^{i(k\sin\theta o - 2\pi/L)x}. \end{aligned} \tag{2.145}$$

The first term represents the field of the incident plane wave attenuated by a factor *a* and represents an *un-diffracted* beam. The two other terms represents the fields of plane waves propagating in two different directions given by

$$\begin{aligned} k\sin\theta_o + (2\pi/L) &= k\sin\theta_{+1} \\ k\sin\theta_o - (2\pi/L) &= k\sin\theta_{-1} \end{aligned}. \tag{2.146}$$

These two waves are the *first order diffracted* waves. The two relations (2.146) can be written as

$$\sin\theta_{\pm 1} - \sin\theta_o = \pm\lambda/L \tag{2.147}$$

that is a relation which gives the angle formed by the diffracted beams as a function of the incidence angle θ_o, period of the grating L and wavelength λ. To have the diffracted field onto a screen one must propagate these fields coming out from the grating. We are not interested in this. Rather we want to see what happens when the period of the grating decreases.

For the sake of simplicity let us assume that the incoming wave is normal to the grating, i.e. $\theta_o = 0$. From (2.147) it is

$$\sin\theta_{\pm 1} = \pm\lambda/L. \tag{2.148}$$

Until $L > \lambda$, this equation has solutions for $\theta_{\pm 1}$. When $L = \lambda$, $\sin\theta_{\pm 1} = \pm 1$ which means that the diffracted beams are grazing the grating surface. However, when $L < \lambda$ there are no real solutions for $\sin\theta_{\pm 1}$ and evanescent waves are produced. $L = \lambda$ is therefore the maximum period of the grating that gives homogeneous diffracted waves. This result shows in another way what we said in the preceding section on resolution because it states that all details represented by Fourier components of spatial frequency less than λ originates evanescent waves. Their contribution to an image in the far field is lost and cannot be recovered in any way.

Chapter 3
Evanescent Waves in Optical Waveguides

3.1 Introduction

The simplest optical waveguide is a dielectric slab. Because of their simple geometry, guided and radiation modes of slab waveguides can be described by simple mathematical expressions. The study of slab waveguides and their properties is thus useful in gaining an understanding of the waveguiding properties of more complicated dielectric waveguides. In many applications optical fibres are used, and we will describe also this kind of structure. In some cases more than two layers of different refractive index are used. These are *photonic band gaps* (PBGs), that also are discussed. Outside the guiding region, evanescent waves are present that are fundamental to describe many of the properties of waveguides, and we will stress this aspect here. Using evanescent waves allows to design and build a number of sensors with planar and fibre geometries, that are briefly examined at the end.

3.2 Planar Waveguides

A dielectric slab waveguide[1] is shown schematically in Fig. 3.1. It consists of a high-index dielectric layer surrounded on either sides by lower index materials. The slab is infinite in extent in the yz plane and finite in the x-direction. The core region of the waveguide is assumed to have refractive index n_2 and is deposited on a substrate with refractive index n_1. The refractive index of the medium above the core – usually called *cladding* - is indicated as n_3.

In order to achieve true mode guidance it is necessary that $n_2 > n_1, n_3$.

If $n_1 = n_3$ the guide is called a *symmetric* slab waveguide. In case that $n_1 \# n_3$, the slab waveguide is said to be *asymmetric*.

[1]There is a number of excellent books on waveguides. See, for example, Hunsperger RG (1982) Integrated optics theory and technology. Springer, Berlin

© Springer International Publishing AG 2017

M. Bertolotti et al., *Evanescent Waves in Optics*, Springer Series in Optical Sciences,
DOI 10.1007/978-3-319-61261-4_3

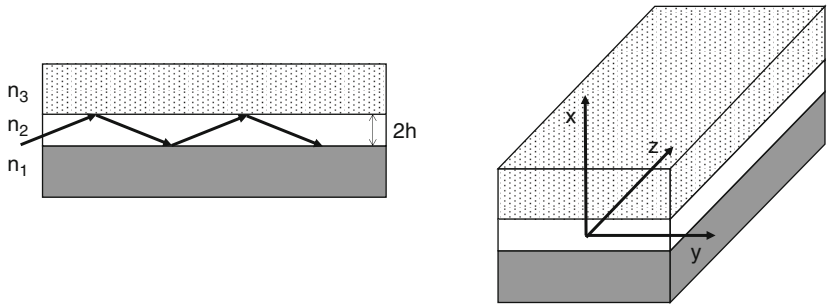
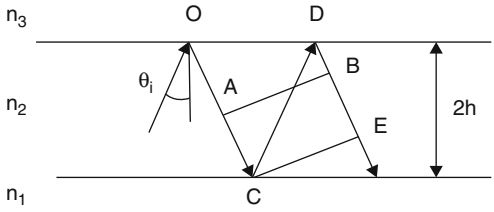


Fig. 3.1 A sketch of a planar waveguide with a propagating ray inside (left) and the 3-D representation (right)

Fig. 3.2 The ray representation in a planar waveguide



A naïve but simple way to understand propagation in the structure is to imagine that light rays are propagating into the waveguide. When they hit the interface with the cladding and substrate material with an angle larger than the critical angle they are totally reflected into the structure and are consequently trapped inside it (Fig. 3.2). To be more specific, let us consider a ray that hit the boundary planes of the waveguide with an angle θ_i. To be totally reflected the angle θ_i should satisfy the following relations

$$n_2 \sin\theta_i > n_3 \quad ; \quad n_2 \sin\theta_i > n_1 . \tag{3.1}$$

We note that the plane wave associated to the ray has a propagation vector whose component along z is $n_2 k \sin\theta$. If we wish that a field propagates in the structure along z, its dependence from z should be of the form $\exp(i\beta z)$ and therefore

$$\beta = n_2 k \sin\theta . \tag{3.2}$$

To find the possible values for β it suffices to find the θ values which allow the field propagation. With reference to Fig. 3.2 let us consider a wavefront that in A is normal to the ray OC. After the reflections in C and D, the radiation in B must have the same phase that in A, belonging to the same wavefront AB. If this was not so, destructive interference would destroy the propagation. The geometric path ACDB = 4 hcos θ, where 2 h is the thickness of the guiding layer, multiplied by $n_2 k$ gives the phase shift due to the path difference. To this phase one should add

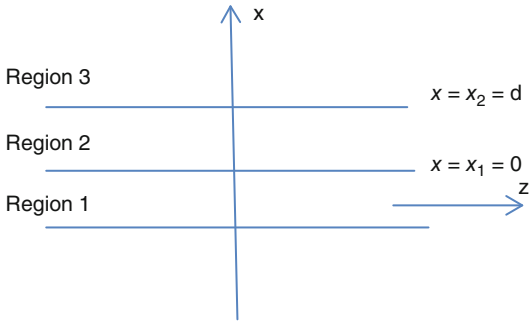


Fig. 3.3 Sketch of the waveguide. The guiding region is Region 2

the phase shifts Φ_C and Φ_D due to total reflection in C and D. The condition which imposes the phase equality in A and B is, therefore

$$4n_2 kh\cos\theta + \Phi_C + \Phi_D = 2m\pi \quad (m = 0, 1, \ldots). \tag{3.3}$$

Solving this equation gives the θ values which may satisfy the propagation requirements.

Full understanding of the problem can be gained solving the pertinent Maxwell's equations (1.1c) and (1.1d), assuming no free charges or currents. Let us consider Fig. 3.3 and assume $n_2 > n_1, n_3$. We consider waves propagating in the direction z, so that $\partial/\partial y = 0$ and look for a solution of the kind

$$\begin{aligned} \mathbf{E} &= \mathbf{E}_o(x)\exp[i(\omega t - \beta z)] \\ \mathbf{B} &= \mathbf{B}_o(x)\exp[i(\omega t - \beta z)] \end{aligned}. \tag{3.4}$$

Equations (1.1c,d) become

$$-\beta E_{oy} = \omega B_{ox} \tag{3.5a}$$

$$\partial E_{oz}/\partial x + i\beta E_{ox} = i\omega B_{oy} \tag{3.5b}$$

$$\partial E_{oy}/\partial x = -i\omega B_{oz} \tag{3.5c}$$

$$\beta B_{oy} = \omega\mu\varepsilon E_{ox} \tag{3.5d}$$

$$-\partial B_{oz}/\partial x - i\beta B_{ox} = i\omega\mu_o\varepsilon E_{oy} \tag{3.5e}$$

$$\partial B_{oy}/\partial x = i\omega\mu_o\varepsilon E_{oz}. \tag{3.5f}$$

Simple inspection shows that the set of (3.5) can be split into two separate subsets; (3.5a)(3.5c) and (3.5e) define a TE mode solution with the electric field transverse to the propagation direction, while (3.5b)(3.5d) and (3.5f) define TM modes with transverse B field.

3.2.1 TE Modes

Let us consider the TE modes, first. From (3.5a) it is

$$B_{ox} = -(\beta/\omega)E_{oy} \tag{3.6}$$

and from (3.5c)

$$B_{oz} = (i/\omega)\partial E_{oy}/\partial x. \tag{3.7}$$

By substituting (3.7) into (3.5e), taking into account (3.5c), one has

$$\partial^2 E_{oy}/\partial x^2 + (\omega^2\mu\varepsilon - \beta^2)E_{oy} = 0, \tag{3.8}$$

that is a wave equation for the component E_{oy}. We may also write

$$\omega^2\mu\varepsilon - \beta^2 = k_o^2 n^2 - \beta^2 \tag{3.9}$$

being $k_o = 2\pi/\lambda_o$ with λ_o, vacuum wavelength. The problem to find the TE modes in the guide is therefore very simple; one has to search for the solutions of the wave (3.8) by applying the boundary conditions (continuity of the tangential component of E and H or of the first derivative of E) at the two interfaces $x = x_1 = 0$ and $x = x_2 = d$. Moreover the two fields E and B must go to zero at infinity.

Equation (3.8) in the three regions is

$$\partial^2 E_{oy}/\partial x^2 + (k_o^2 n_1^2 - \beta^2)E_{oy} = 0 \qquad \text{for } x < 0 \tag{3.10a}$$

$$\partial^2 E_{oy}/\partial x^2 + (k_o^2 n_2^2 - \beta^2)E_{oy} = 0 \qquad \text{for } 0 < x < d \tag{3.10b}$$

$$\partial^2 E_{oy}/\partial x^2 + (k_o^2 n_3^2 - \beta^2)E_{oy} = 0 \qquad \text{for } x > d. \tag{3.10c}$$

The solutions of the equations depend on the sign of the term $(k_o{}^2 n_i - \beta^2)$ $(i = 1,2,3)$. Let us consider (3.10a) and assume

$$(k_o^2 n_1^2 - \beta^2) = q \tag{3.11}$$

positive; the (3.10a) may be written as

$$\partial^2 E_{oy}/\partial x^2 + q\, E_{oy} = 0. \tag{3.12}$$

The solution of an equation of this kind is

$$E_{oy} = A\cos(q)^{1/2}x. \tag{3.13}$$

However this is an oscillating function which cannot vanish for $x \to -\infty$. Therefore the relation (3.11) must be negative. We may write

$$k_o^2 n_1^2 - \beta^2 = -\delta^2. \tag{3.14}$$

Equation (3.10a) can then be written

$$\partial^2 E_{oy}/\partial x^2 - \delta^2 E_{oy} = 0 \qquad \text{for } x < 0 \tag{3.15}$$

which has the obvious solution

$$E_{oy} = Ae^{+\delta x} + A'e^{-\delta x}. \tag{3.16}$$

Because for $x \to -\infty$, E_{oy} must go to zero, it should be $A' = 0$, and therefore

$$E_{oy} = Ae^{\delta x} \qquad x < 0. \tag{3.17}$$

Note that this is the behaviour of an evanescent wave.
Equation (3.10b), putting

$$k^2 = k_o^2 n_2^2 - \beta^2 > 0 \tag{3.18}$$

becomes

$$\partial^2 E_{oy}/\partial x^2 + k^2 E_{oy} = 0 \qquad 0 < x < d \tag{3.19}$$

with solution

$$E_{oy} = B'\cos(kx) + B\sin(kx) \qquad 0 < x < d. \tag{3.20}$$

In the same way, putting

$$-\gamma^2 = k_o^2 n_3^2 - \beta^2 \tag{3.21}$$

Equation (3.10c) becomes

$$\partial^2 E_{oy}/\partial x^2 - \gamma^2 E_{oy} = 0 \qquad x > 0 \tag{3.22}$$

with solution

$$E_{oy} = Ce^{-\gamma x} \tag{3.23}$$

taking into account that $E_{oy} \to 0$ for $x \to \infty$. Again (3.23) represents an evanescent field.

The boundary conditions require that E_{oy} be continuous in x = 0 and x = d. Therefore

$$\begin{array}{ll} A = B' & \text{for } x = 0 \\ B'\cos(kd) + B\sin(kd) = Ce^{-\gamma d} & \text{for } x = d \end{array}. \tag{3.24}$$

That is

$$C = [A\cos(kd) + B\sin(kd)]e^{\gamma d}. \tag{3.25}$$

We may therefore write

$$E_{oy} = Ae^{+\delta x} \qquad \text{for } x < 0 \tag{3.26a}$$

$$E_{oy} = A\cos(kx) + B\sin(kx) \qquad 0 < x < d \tag{3.26b}$$

$$E_{oy} = [A\cos(kd) + B\sin(kd)]e^{\gamma(d-x)} \qquad x > d. \tag{3.26c}$$

For the magnetic induction B we have

$$B_{oz} = (i/\omega)A\delta e^{\delta x} \qquad \text{for } x < 0 \tag{3.27a}$$

$$B_{oz} = (i/\omega)k[B\cos(kx) - A\sin(kx)] \qquad 0 < x < d \tag{3.27b}$$

$$B_{oz} = -(i\gamma/\omega)[A\cos(kd) + B\sin(kd)]e^{\gamma(d-x)} \qquad x > d. \tag{3.27c}$$

The request of continuity for H_{oz} at x = 0 and x = d gives

$$\begin{array}{c} (i/\omega)A\delta = (i/\omega)kB \\ (i/\omega)k[B\cos(kd) - A\sin(kd)] = -(i\gamma/\omega)[A\cos(kd) - B\sin(kd)] \end{array}. \tag{3.28}$$

That is

$$\begin{array}{c} \delta A - kB = 0 \\ -[k\sin(kd) - \gamma\cos(kd)]A + [k\cos(kd) + \gamma\sin(kd)]B = 0 \end{array}. \tag{3.29}$$

This system of homogeneous equations has solutions only if the determinant of the system vanishes, that is

$$\delta[k\cos(kd) + \gamma\sin(kd)] + k[-k\sin(kd) + \gamma\cos(kd)] = 0. \tag{3.30}$$

In (3.30) all quantities are known, with the exception of β that determines the way the wave propagates and is called *propagation constant*. The Equation (3.30) may be written also as,

$$(k\gamma + k\delta)\cos(kd) + (\gamma\delta - k^2)\sin(kd) = 0 \tag{3.31}$$

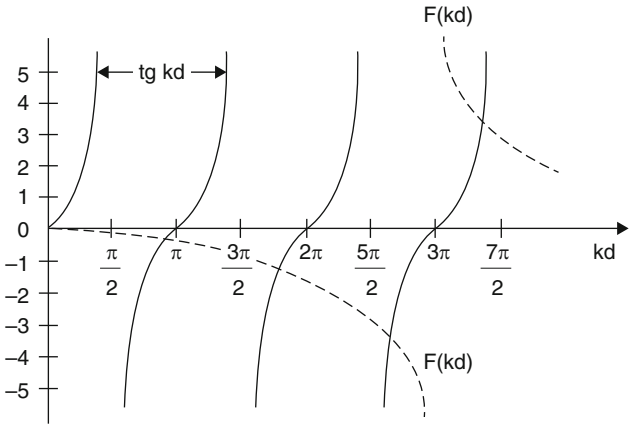


Fig. 3.4 The solid lines are the branches of the tangent as a function of kd. The dashed lines represent the function F(kd). The crossing points of the solid and dashed lines correspond to solutions

or

$$\mathrm{tg}(kd) = k(\gamma + \delta)/(-k^2 + \gamma\delta). \tag{3.32}$$

The equation can be solved numerically or graphically. A graphical solution[2] is shown in Fig. 3.4 using the function F(kd) defined as

$$\begin{aligned} F(kd) &= k(\gamma + \delta)/(-k^2 + \gamma\delta) = kd[(\gamma + \delta)/d]/\left[(-k^2 d^2/d^2) + \gamma\delta\right] \\ &= kd(\gamma + \delta)d/\left[-(kd)^2 + \gamma\delta d^2\right]. \end{aligned} \tag{3.33}$$

Because

$$k^2 = n_2^2 k_o^2 - \beta^2 \tag{3.34}$$

one has

$$\beta^2 = n_2^2 k_o^2 - k^2 \tag{3.35}$$

that substituted into (3.21) and (3.14) allows to write

$$\begin{aligned} \gamma^2 &= (n_2^2 - n_3^2)k_o^2 - k^2 \quad \text{and} \quad \delta^2 = (n_2^2 - n_1^2)k_o^2 - k^2 \\ V^2 &= (n_2^2 - n_3^2)k_o^2 d^2 \qquad\qquad Z^2 = (n_2^2 - n_1^2)k_o^2 d^2 \end{aligned}. \tag{3.36}$$

[2]Marcuse D (1974) Theory of dielectric optical waveguide. Academic Press, London

The function F(kd) can then be written in a more compact form as

$$F(kd)=kd\left\{\left[V^2-(kd)^2\right]^{1/2}+\left[Z^2-(kd)^2\right]^{1/2}\right\}\Big/\left|\left|(kd)^2-\left\{\left[V^2-(kd)^2\right]\left[Z^2-(kd)^2\right]\right\}^{1/2}\right|\right|. \tag{3.37}$$

Figure 3.4 is for V = 11 and Z = 24, and gives tg(kd) and F(kd) as a function of kd. The poles of the curve F(kd) occur at the points where the denominator of (3.37) vanishes. The curve F(kd) ends at the point

$$V = kd \tag{3.38}$$

since one of the square roots of (3.37) becomes imaginary as kd exceeds the value given by (3.38).

The kd coordinates of the crossing points of the solid and dashed curves, represent solutions of (3.32). Each solution gives one TE mode of the slab waveguide.

For the numerical values used in the figure, in the considered guide 4 propagation modes are allowed. As the value of V decreases, the endpoint of F(kd) moves to the left, so that it crosses fewer branches of the curve tg(kd). Therefore the number of allowed modes decreases. If V becomes small enough, the end point of F(kd) moves to the left of the first branch of tg(kd) and no guided mode may exists. This condition is called *cut-off*. At cut-off $\beta = n_3 k$. However, if the guide is symmetric ($n_1 = n_3$) when (3.38) occurs, F(kd) = 0, that is the end point of F(kd) is on the axis kd and at least one guided mode always exists.

If $n_2 = n_3$, one has $\gamma = \delta$ and (3.32) reduces to

$$tg(2k_o d/2) = (2\gamma/k_o)/[1-(\gamma/k_o)]^2 = [2tg(k_o d/2)]/\left[1 - tg^2(k_o d/2)\right] \tag{3.39}$$

that is a second-order equation in $tg(k_o d/2)$ with the solutions

$$tg(k_o d/2) = \gamma/k_o \quad \text{and} \quad tg(k_o d/2) = -k_o/\gamma. \tag{3.40}$$

The first of (3.40) describes the so called even modes and the second one the odd modes. The fact that even and odd modes result can be verified from (3.26b) by shifting the coordinate origin to the centre of the core of the symmetric slab. Even modes have the electric field spatial distribution that is symmetric to it, while odd modes are anti-symmetric.

3.2.2 TM Modes

TM modes have the B_y component transverse to the direction of motion and the two components of the electric field E_x and E_z.

Proceeding in an similar way as for TE modes one has

$$\partial E_{oz}/\partial x + i\beta E_{ox} = i\omega B_{oy} \tag{3.41a}$$

$$\beta B_{oy} = \omega\mu\varepsilon E_{ox} \tag{3.41b}$$

$$\partial B_{oy}/\partial x = i\omega\mu_o\varepsilon E_{oz} \tag{3.41c}$$

from which

$$E_{ox} = (\beta/\omega\mu\varepsilon)B_{oy} \tag{3.42a}$$

$$E_{oz} = -(i/\omega\mu\varepsilon)\partial B_{oy}/\partial x \tag{3.42b}$$

and eventually

$$\partial^2 B_{oy}/\partial x^2 + (\omega^2\mu\varepsilon - \beta^2)B_{oy} = 0. \tag{3.43}$$

Boundary conditions require that E_{ox}, $H_{oy} = B_{oy}/\mu_o$ be continuous at the interfaces $x = 0$ and $x = d$. The solutions of (3.41) which satisfy these conditions and vanishes for $x = \pm\infty$ are

$$\begin{aligned} B_{oy} &= (\beta/|\beta|)Ce^{+\delta x} && \text{for } x < 0 \\ B_{oy} &= (\beta/|\beta|)[C\cos(kx) + D\sin(kx)] && \text{for } 0 < x < d\,. \\ B_{oy} &= (\beta/|\beta|)Ce^{k(d-x)} && \text{for } x > d \end{aligned} \tag{3.44}$$

The term $\beta/|\beta|$ is made explicit to assure that the transverse magnetic field changes its sign when the direction of propagation is inverted.

The components E_{oz} may be derived, and requiring continuity at $x = 0$ and $x = d$ one obtains the eigenvalue equation

$$tg(kd) = n_1^2 k(n_1^2\gamma + n_2^2\delta)/\left[n_2^2 n_3^2 k^2 - n_1^4\gamma\delta\right]. \tag{3.45}$$

The solutions of (3.45) are obtained by a graphic substantially similar to Fig. 3.4 For the special case of symmetric waveguide ($n_1 = n_3$) for even modes one has

$$tg(k_o d/2) = (n_1/n_2)^2\gamma/k_o \tag{3.46}$$

and for odd modes

$$tg(k_o d/2) = -(n_2/n_1)^2 k_o/\gamma. \tag{3.47}$$

The coefficients A and C are obtained by integrating the z component of the power flow vector (Poynting vector)

$$S_z = (\tfrac{1}{2})\mathrm{Re}(\mathbf{E}\mathbf{x}\mathbf{H}^*)\cdot u_{\mathbf{z}} \tag{3.48}$$

over the infinite transverse cross section of the waveguide. For the slab with its infinite extension in the y direction one obtains the power P_l per unit length ($P_l = P/l$) in the y direction. We eventually obtain, for a TM mode

$$A^2 = 4\pi^2 \omega \mu P_l / |\beta| \left[d + (1/\gamma) + (1/\delta) \right] (k^2 + \delta^2). \tag{3.49}$$

3.2.3 The Radiation and Substrate Modes

A slab waveguide can support additional modes besides the guided ones, as shown in Fig. 3.5. If we consider the nature of solutions as a function of the propagation constant β at a fixed frequency ω, for $\beta > k_o n_2$ (the regime depicted by the first curve from the right) it follows directly from (3.10) that E(x) is exponential in all three layers of the waveguide and the continuity conditions give the curve shown in the figure. The field increases without bound away from the waveguide so that the solution is not physically realizable and thus does not correspond to a real wave.

For $k_o n_3 < \beta < k_o n_2$ one has the guided modes already described. This situation corresponds to the second and third curve in Fig. 3.5 from the right. Mode solutions for $k_o n_1 < \beta < k_o n_3$ correspond to exponential behaviour in region 1 and to

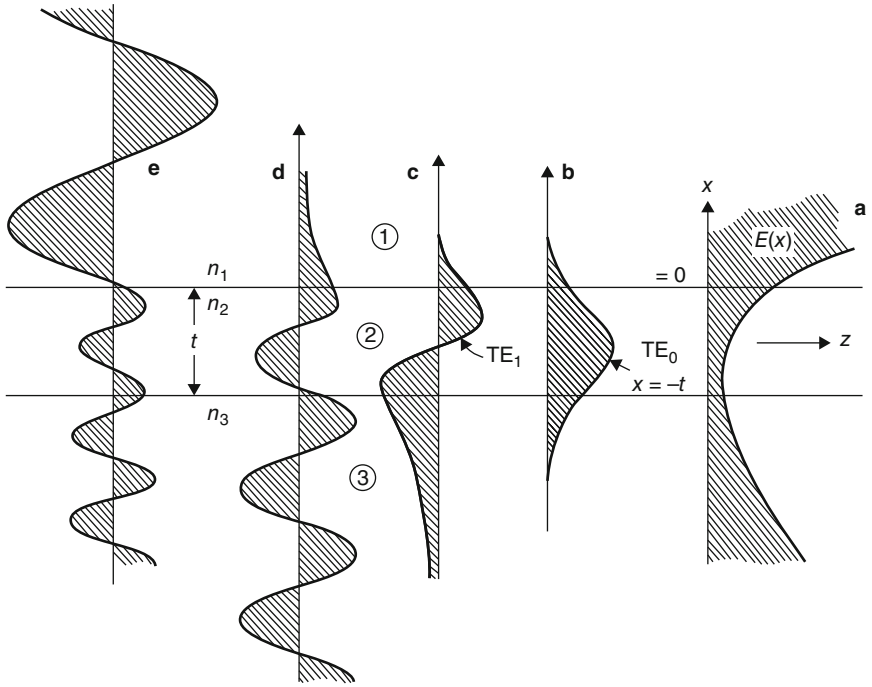


Fig. 3.5 The different regimes of the propagation constant of a slab waveguide. (Modified from A Yariv, Quantum Electronics, Wiley 1975)

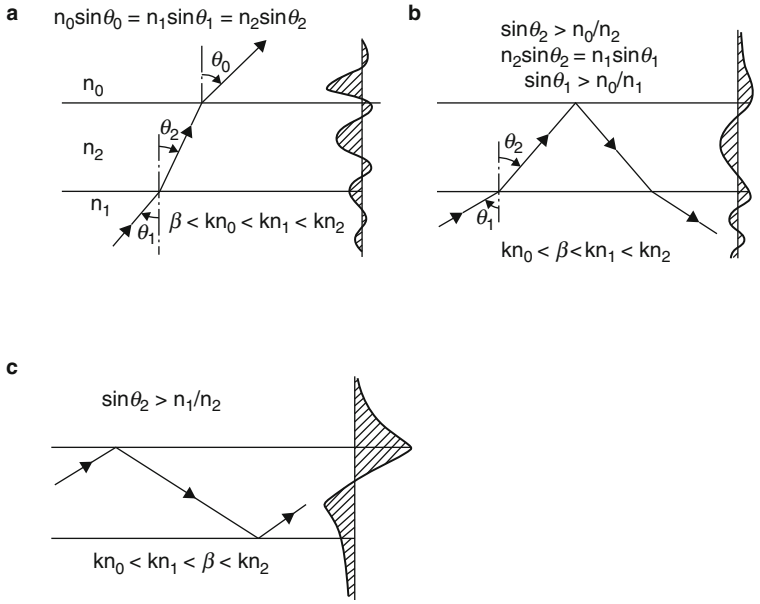


Fig. 3.6 The geometrical interpretation of the various waveforms characterizing the modes. (**a**) shows radiation modes in air; (**b**) substrate radiation modes, and (**c**) guided modes

sinusoidal behaviour in regions 2 and 3 (the fourth curve in Fig. 3.5); we will refer to these modes as *substrate radiation modes*. For $0 < \beta < k_o n_1$, as the last curve shown, the solution for E(x) becomes sinusoidal in all three regions. These are the so-called *radiation modes* of the waveguide. In the last two regimes β is a continuous variable.

Figure 3.6 shows, from the point of view of geometrical optics what happens to the ray for modes whose wave forms are shown in Fig. 3.5.

A different plot to describe modes in a waveguide is the one in which the propagation constant β is plotted as a function of frequency ω (Fig. 3.7), giving the dispersion relation β(ω). In the figure, the first three modes are shown. At the cut-off frequency, the propagation constants assume the value of the lower bound $n_1 k$ and as ω (or the thickness h) increases, β approaches its upper bound $n_3 k$ and more and more guided modes exist. In addition to the discrete spectrum of the guided modes, the diagram also shows the continuous spectrum of the radiation modes.

All modes are orthogonal. For guided modes, orthogonality requires that

$$\int_{\text{area}} \left[\mathbf{E_i} \mathrm{x} \mathbf{H_j}\right] \cdot \mathrm{dA} = \delta_{ij} \mathrm{S} \tag{3.50}$$

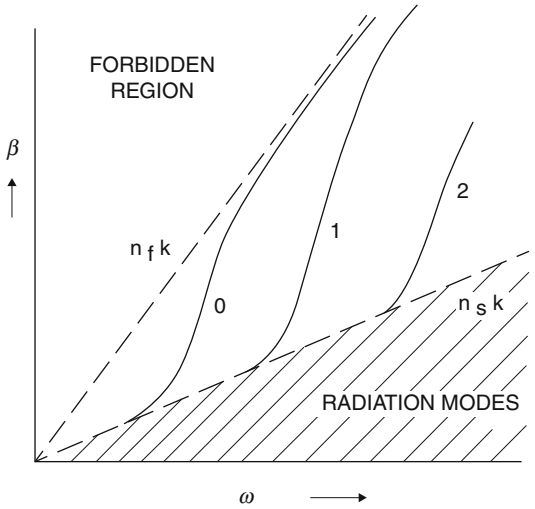


Fig. 3.7 Typical dispersion plot for a planar waveguide

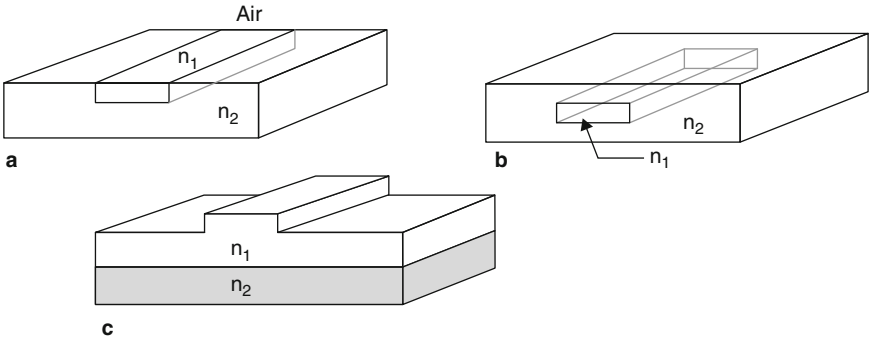


Fig. 3.8 some configurations for rectangular waveguides. In (**a**) the lateral index is the same of the cladding (*surface waveguide*); in (**b**) the guiding layer is completely surrounded by the cladding layer (*buried waveguide*); in (**c**) the lateral confinement is established by the dielectric ridge on top of the substrate (*ridge waveguide*)

where δ_{ij} is the Kroenecker delta function. The area of the integral is the infinite xy plane at a particular value of z, and E and H represent normalized modes of the system. The modes of a given system form a complete set, that is the allowed modes span the entire space of the system.

Optical waveguides which provide two-dimensional confinement of the field are constructed from slab waveguides decreasing laterally the index of the slab in the y direction in various ways (Fig. 3.8). The field of a mode is concentrated in the region of highest index. It decays into the substrate and into the region of decreased index. The mode pattern is still roughly that of the slab guide except for its finite extent in the y direction. The modes are still called TE and TM, even

though they do not possess strictly transverse electric or magnetic fields. There is not any general analytical solution for these structures. Their analysis can be achieved by using numerical methods.

3.3 Coupling of Light to a Planar Waveguide

To excite the modes in a planar waveguide one may use either excitation at one end of the film or excitation through the surface.[3]

When possible, light can be coupled directly to a waveguide with a lens simply focusing at the edge of the waveguide and shaping the beam profile so that it matches to the modal distribution in the waveguide (Fig. 3.9). This kind of coupling is called *butt coupling* or *end-fire coupling*. However it requires proper mechanical polishing and gluing for efficient coupling. Moreover, when the film thickness is of the order of the wavelength of light, direct coupling of a collimated beam at the edge of the guide is inefficient because of the geometric mismatch between the beam and the guide.

Coupling directly a radiation mode outside the surface of the waveguide to a guided mode inside the waveguide through its surface presents a severe difficulty. For the sake of simplicity let us consider a guide in air, so that $n_3 = n_o = 1$. The external field will surely extend into the waveguide, either as an evanescent field, or as a travelling wave passing through the structure; however the phase velocities of the external wave and of the mode are different. The radiation field propagates with a spatial wave-vector k_o. The field in the waveguide propagates with a spatial wave-vector $\beta \sim k_o n_2$. Therefore the two waves propagate with different velocities and the radiation fields rapidly get out of phase cancelling themselves.

For a mode of a waveguide with refractive index n_2 inside, n_o outside and n_3 for the substrate, the propagation constant β along the z direction has to satisfy the

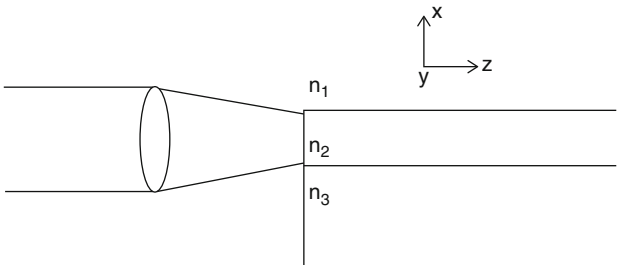


Fig. 3.9 End-fire coupling of a planar waveguide

[3]A simple discussion of the characterization methods of integrated waveguides can be found in Pelli S, Righini GC (1994) In: Martellucci S, Chester AN, Bertolotti M (eds) Advances in integrated optics. Plenum Press, New York, p 1

conditions $k_o n_o < \beta < k_o n_2$ and $k_o n_3 < \beta < k_o n_2$. For an incident plane wave from outside, its horizontal wave-number is $k_z = k_o n_o \sin\theta < k_o n_o$ for any real incident angle θ. Thus, the wave cannot be phase matched to the guide mode which always has $\beta > k_o n_o$.

However, the excitation of the guided mode can be obtained in surface waveguides in which cladding is air using a prism[4] (Fig. 3.10). In this coupler, the prism (refractive index n_p) is placed above the guide and is separated from it by a small gap of low refractive index ($n_p > n_o, n_3$). The incident light is totally reflected at the base of the prism, and the waves in the prism and in the waveguide are coupled through their evanescent fields in the gap. The coupler permits excitation of any one of the film modes by proper orientation of the direction of the incident beam (Fig. 3.10).

To better understand the coupling mechanism,[5] let us consider the optical structure in Fig. 3.11, where a prism is located a distance s from the surface of the waveguide. The prism has an index of refraction n_p, while the waveguide on the surface has a guiding film index n_2 on a substrate with index n_1.The medium between the prism and the waveguide is air of index $n_3 = n_o = 1$ (in general it could be any transparent material with $s < \lambda$).

The incident beam is directed into the prism at an angle such that total internal reflection occurs at the n_p-air interface. This is satisfied if the angle of incidence θ is greater than the critical angle

$$\theta > \theta_c = \sin^{-1}(1/n_p) \tag{3.51}$$

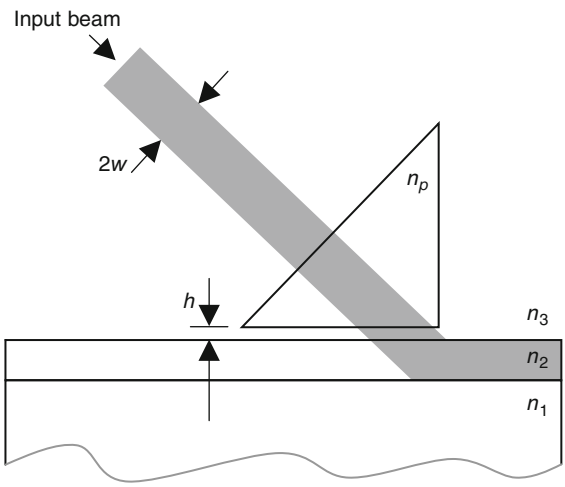


Fig. 3.10 Disposition for prism coupling

[4]Tien PK (1971) Appl Opt. 10:2395

[5]The theory is in Tien PK, Ulrich R (1970) J Opt Soc America 60:1325

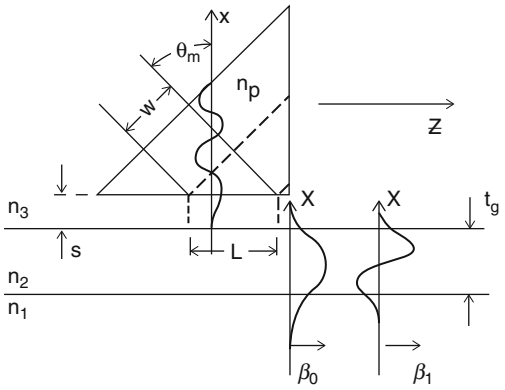


Fig. 3.11 Prism coupling. The electric field distribution of the prism mode and the two first modes (m = 0,1) in the waveguide are shown. From Hunsperger 1984[6]

Inside the prism, the incident and the reflected waves form a standing wave pattern. The k-vector for the field in the prism can be described in terms of its components

$$\mathbf{k_o} n_p = n_p(\pm k_o \cos\theta \mathbf{u_x} + k_o \sin\theta \mathbf{u_z}) \tag{3.52}$$

where $\mathbf{u_x}$ and $\mathbf{u_z}$ are the unit vectors in the x and z direction, respectively.

Below the prism-air interface, the field decays exponentially with increasing distance.

If the prism spacing s is small enough so that the tail of the waveguide modes overlap the tail of the prism mode, there may occur coherent coupling of energy from the prism mode.

The x-component of the propagation coefficient is imaginary in this region, but the z-component remains the same as inside the prism. Since the z-component of k depends on the angle of incidence, it is possible to adjust the angle so that the waves travel at the same velocity as those in the waveguide. When this happens, strong coupling occurs. Specifically to maximize the interaction between the fields in the prism and the waveguide, the angle of incidence of the beam with respect to the bottom surface should be

$$k_o n_p \sin\theta = \beta. \tag{3.53}$$

This is called the *phase-matching condition*.

Note that this angle is measured inside the prism. The external angle at which the beam enters the prism must be adjusted to account for refraction at the prism-air interface to satisfy this equation. The utility of prism coupling is that the angle of incidence can be adjusted to satisfy (3.53) for each and every mode in the waveguide (not simultaneously however) allowing the selective coupling of energy to individual modes, and allowing the experimental determination of mode structure.

[6]Hunsperger RG (1984) Integrated optics: theory and technology, 2nd edn. Springer, Berlin

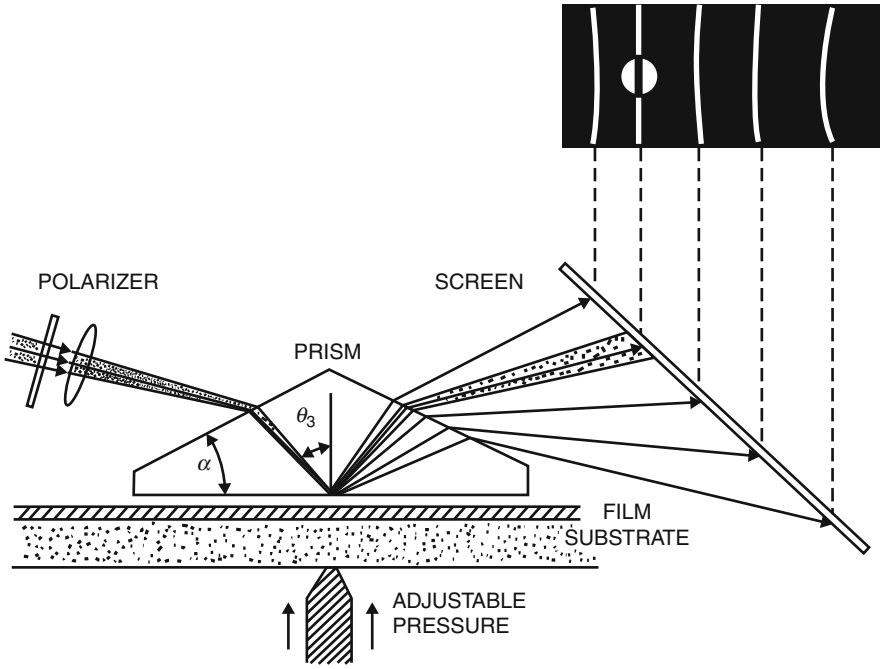


Fig. 3.12 Coupling in and out with a prism.
(From P K Tien and R Ulrich, J Opt. Soc. America 60, 1325 (1970) © Optical Society America)

Because of the size of the prism, the interaction between prism and waveguide modes can occur only over the length L. The theory of weakly coupled modes (&3.4) indicates that a complete interchange of energy between phase-matched modes occurs if the interaction length in the z direction satisfies the relation

$$KL = \pi/2 \tag{3.54}$$

where K is the coupling coefficient. This condition for complete coupling assumes that the amplitude of the electric field is uniform over the entire width W of the beam. In a practical case this is never true. For a Gaussian beam shape, the maximum coupling efficiency is about 80%. Also note that in order to get 100% coupling (for a plane wave) the trailing edge of the beam must exactly intersect the right-angle corner of the prism. If it intersects too far to the right, some of the incident power will be either reflected or transmitted directly into the waveguide and will not enter the prism mode. If the beam is incident too far to the left, some of the power coupled into the waveguide will be coupled back out into the prism. This is the case of Fig. 3.12 in which a 45° prism is used.

Because of film inhomogeneities, the optical energy in the excited mode is rapidly scattered into the other modes of the film and is then coupled back to the

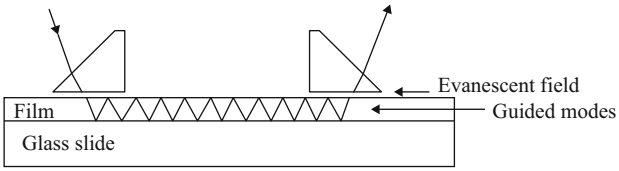


Fig. 3.13 Coupling light in and out from a planar waveguide

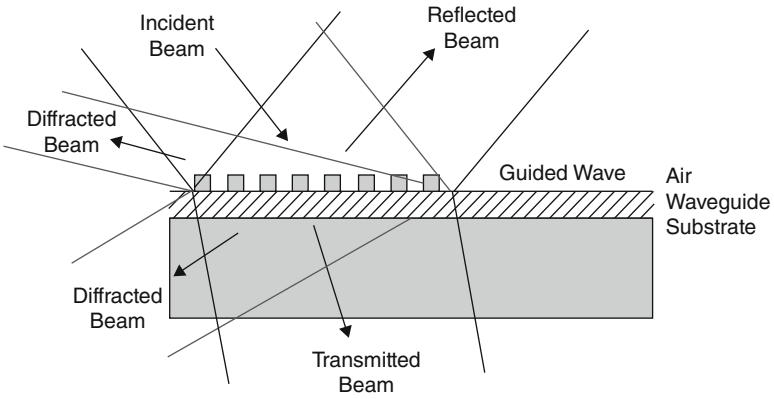


Fig. 3.14 A grating coupler

outside medium by the same prism. One thus see on a screen a series of bright lines (the so called *m-lines*) with a bright spot on one of these lines. The bright spot is the beam totally reflected from the base of the prism. Inside this spot there is a dark line which represents the energy lost because of the coupling to the excited modes. Each line represents a mode of a different order m (from which the name). In this way it is possible to assess the modes which propagate in a waveguide.

The theory of coupling has been discussed by various authors.[7]

Once a mode is excited, a prism could also be used to decouple the radiation from the waveguide to the outside (Fig. 3.13).

Another method widely used to couple radiation into a guide is to use a grating (Fig. 3.14). In this case an incident beam strikes a grating that is in contact with the guide and is diffracted according to the grating's theory. If the grating structure is oriented along the z direction with period p, the relation between incident and diffracted light wavevector is

$$k_{out} = k_{in} \pm 2m\pi/p \tag{3.55}$$

[7]Tien PK, Ulrich R (1970) J Opt Soc America 60:1325; Chang KC, Tamir T (1980) Appl Opt. 19:282; Harris JH, Winn RK, Dalgoutte DG (1972) Appl Opt. 10:2234

where m is an integer. The grating is therefore able to add or subtract integer units of $2\pi/p$ to the z component of the incident wave. It is then possible to choose an angle θ such that

$$\beta = k_o n_a \sin\theta + 2\pi/p \tag{3.56}$$

so coupling the external field to the mode propagating in the waveguide.[8] A grating could also be used to decouple the radiation from the waveguide, as mentioned for the case of the prism.

3.4 Coupling of Two Waveguides

Modal coupling between optical modes may occur for various reasons in a single waveguide or in near waveguides. In some cases scattering from imperfections can give raise to the coupling. In other cases coupling happens because of the presence of evanescent tails in the modes.[9] We describe here this latter case, considering a very special case of coupling to emphasize the role played by evanescent waves.

If two dielectric waveguides are placed alongside each other, some of the field of one guide reaches the other guide. We may treat the problem as a small perturbation (*weak coupled theory*). The presence of the second guide distorts the field of the guided mode of the first guide. The distorted field can be expressed as a superposition of the mode fields of each guide plus a small additional field. Since the mode field of the second guide is required to express the distortion of the field of the first guide, it is clear the two mode interact. Complete transfer of power is possible. It is required only that the phase velocities of the modes in either guide are identical in the absence of coupling. Modes with different phase velocities cannot interact very effectively, and only very little power can be exchanged.

The exchange of power may be very effective if the two waveguides are very near to each other so to reciprocally influence themselves. Let us consider a double planar waveguide as shown in Fig. 3.15. One may describe what happens assuming that in each guide the same modes are propagating that were found for the isolated guide and that the presence of an evanescent field in the region between the two guides allows the possibility of coupling the propagation mode in a guide to a mode propagating in the nearby wave guide.

The field in the cores of the ideal (uncoupled) waveguides will be of the form

$$E_A(x,y,z) = A\cos(k_x x + \phi_x)\cos\left[k_y(y + (a+b)\right]\exp(-i\beta z) = A\boldsymbol{\varepsilon}(x)\boldsymbol{\varepsilon}(y) \tag{3.57a}$$

$$E_B(x,y,z) = B\cos(k_x x + \phi_x)\cos\left[k_y(y - (a+b)\right]\exp(-i\beta z) = B\boldsymbol{\varepsilon}(x)\boldsymbol{\varepsilon}(y') \tag{3.57b}$$

[8] Dakas ML et al (1970) Appl Phys Lett. 16:523; Kogelnik HK, Sosnowski T (1970) Bell Syst Tec J. 49:1602

[9] Marcuse D (1971) Bell Syst Tec J. 50:1791

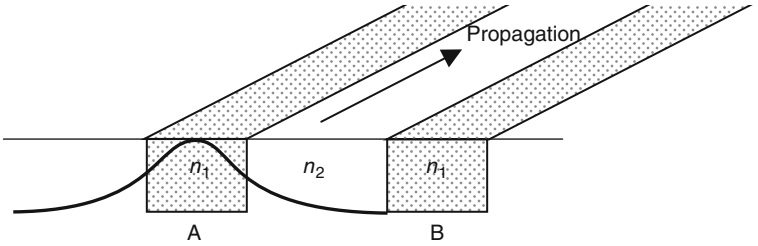


Fig. 3.15 Two rectangular waveguides of refractive index n_1 separated by a region of refractive index n_2 and the field representation of a mode in guide A that with its tail can couple to guide B

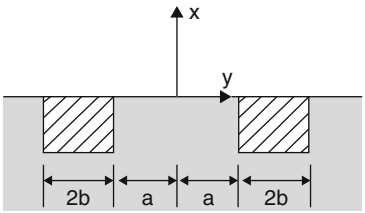


Fig. 3.16 Geometry of the coupled waveguides

where $\boldsymbol{\varepsilon}(x)$, $\boldsymbol{\varepsilon}(y)$ are the mode field distribution in both waveguides that have the same transverse and longitudinal wave-vectors, k_x, k_y, and β, respectively, and the center for each waveguide is a distance a+b from the origin of the coordinate system (Fig. 3.16). Outside the core the field decays exponentially.

We describe the transfer of power from one guide to the other through a perturbation in the polarization P_{pert}, so that D may be written as

$$D = \varepsilon E + P_{pert}. \tag{3.58}$$

The wave equation for E now becomes

$$\nabla^2 E_y = \mu\varepsilon \partial^2 E_y / \partial t^2 + \mu \partial^2 P_{pert} / \partial t^2 \tag{3.59}$$

The polarization perturbation in waveguide B arises from the presence of the evanescent tail of mode A. The perturbation is actually due to the difference in index the evanescent field sees when it is in the core of waveguide B, compared to the normal cladding index. The polarization induced by mode A acts as a source to excite mode B. From mode B's perspective, the perturbation is

$$\begin{aligned} P_{pert}(x,y,z) &= \varepsilon_o \left[n_1^2(x,y) - n_2^2\right] E_A(x,y) \\ &= \varepsilon_o \left[n_1^2(x,y) - n_2^2\right] A\boldsymbol{\varepsilon}(x)\boldsymbol{\varepsilon}(y)\exp(-i\beta z). \end{aligned} \tag{3.60}$$

The general equation of motion for the structure is

$$\begin{aligned} &-(\partial B/\partial z)e^{-i(\beta z-\omega t)}+cc=-(i/2\omega)\partial^2/\partial t^2\int B\boldsymbol{\varepsilon}(x)\boldsymbol{\varepsilon}(y)P_{pert}dxdy \\ &=(i\omega/2)\int \varepsilon_o\left[n^2(x,y)-n_2^2\right]BA\boldsymbol{\varepsilon}^2(x)\boldsymbol{\varepsilon}(y)\boldsymbol{\varepsilon}(y')e^{-(\beta z-\omega t)}\Big]dxdy+c.c. \\ &=iKAe^{-i(\beta z-\omega t)} \end{aligned} \tag{3.61}$$

where the coupling constant K is

$$K=(\omega/4)\int \varepsilon_o\left[n^2(x,y)-n_2^2\right]B\boldsymbol{\varepsilon}^2(x)\boldsymbol{\varepsilon}(y)\boldsymbol{\varepsilon}(y')dxdy \tag{3.62}$$

and A and B are normalized amplitudes. A similar equation to (3.61) can be written for the influence that B has on A, so that we end with two coupled equations

$$dA/dz=-iKB \tag{3.63a}$$

$$dB/dz=+iKA. \tag{3.63b}$$

Taking the derivative of the first equation, and putting it into the second equation gives

$$d^2A/dz^2=-iK^2A \tag{3.64}$$

which can be directly solved. If we assume, for example that the initial conditions for the problem are A = 1, and B = 0, then the solutions become

$$A(z)=\cos(Kz) \tag{3.65a}$$

$$B(z)=-i\sin(Kz). \tag{3.65b}$$

So energy can be transferred from one guide to the other. Complete transfer from guide A to B occur when Kz = π(1/2+q) with q an integer. Therefore the length of the interaction region for total exchange of power from one guide to the other is

$$z_o=\pi/2K+q\pi/K. \tag{3.66}$$

Other lengths produce values between 0 and 100% transfer.

3.5 Optical Fibres

A fiber is a cylindrical structure of two or more layers of different refractive index. The central cylinder called *core* has the higher refractive index. Light is guided in this region. It is surrounded by a *cladding* of slightly lower refractive index which

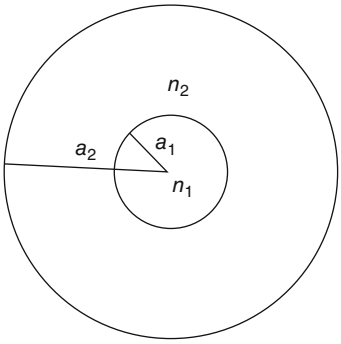


Fig. 3.17 Cross section of a typical optical fibre. The protecting layer is not shown

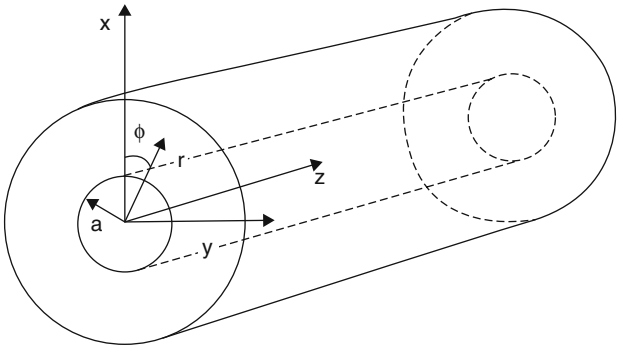


Fig. 3.18 Cylindrical coordinates to describe propagation in a fibre

is finally surrounded by a third cylindrical layer usually made in plastic to protect the fiber from environmental hazards and abrasions (Fig. 3.17).

The refractive index in the core may be constant along the radius and have a discontinuity when passing in the cladding, in this case the fiber is said to be of the *step index* type. Another common kind of fiber is the one with a core refractive index which changes along the radius with a prescribed law. For example according to a parabolic law. This kind of fiber is called *graded index* fiber.

The determination of the propagating modes in a fibre is somehow more complex than for a slab guide because of the cylindrical symmetry of the system.

The use of cylindrical coordinates to treat this case is to be preferred (Fig. 3.18). The electric field is

$$\mathbf{E}(\mathrm{r},\phi,\mathrm{z}) = \mathbf{u_r}\mathrm{E_r}(\mathrm{r},\phi,\mathrm{z}) + \mathbf{u_\phi}\mathrm{E}_\phi(\mathrm{r},\phi,\mathrm{z}) + \mathbf{u_z}\mathrm{E}_z(\mathrm{r},\phi,\mathrm{z}). \tag{3.67}$$

The wave equation is

$$\nabla(\nabla\cdot\mathbf{E}) - \nabla\mathrm{x}\nabla\mathrm{x}\mathbf{E} - \mu\varepsilon(\partial^2\mathbf{E}/\partial\mathrm{t}^2) = 0. \tag{3.68}$$

This equation cannot be decomposed into the three components because the components E_r and E_ϕ are coupled and change with propagation. The component E_z can instead be found from (3.68). It is

$$\partial^2 E_z/\partial\rho^2 + (1/\rho)\partial E_z/\partial\rho + (1/\rho^2)\partial^2 E_z/\partial\phi^2 + \partial^2 E_z/\partial z^2 + n^2 k_o^2 E_z = 0 \tag{3.69}$$

where n is a prescribed function of ρ according to the kind of fiber.

Equation (3.69) is easily solvable by separation of variables

$$E_z(\rho, \phi, z) = F(\rho)\Phi(\phi)Z(z). \tag{3.70}$$

By substituting (3.70) into (3.69) we have

$$d^2Z/dz^2 + \beta^2 Z = 0 \tag{3.71a}$$

$$d^2\Phi/d\phi^2 + m^2\Phi = 0 \tag{3.71b}$$

$$d^2F/d\rho^2 + (1/\rho)dF/d\rho + \left[n^2k_o^2 - \beta^2 - (m/\rho)^2\right]F = 0. \tag{3.71c}$$

The solution of (3.71a) is simply

$$Z = \exp(i\beta z) \tag{3.72}$$

where β is the propagation constant.

The solution of (3.71b) is

$$\Phi = \exp(im\phi). \tag{3.73}$$

Here the constant m is limited to take only integer values because the field must be periodic in ϕ with a period 2π.

Equation (3.71c) is a differential equation satisfied by Bessel functions. Its general solution for a step index fibre may be written as

$$F(\rho) = \begin{cases} AJ_m(p\rho) + A'Y_m(p\rho) \\ CK_m(q\rho) + C'I_m(q\rho) \end{cases} \quad \begin{matrix} \rho < a \\ \rho > a \end{matrix} \tag{3.74}$$

where A, A′, C and C′ are constants, and J_m and Y_m are Bessel functions of first and second species, respectively and K_m and I_m are Hankel modified functions (of first and second species, respectively). They are shown in the Fig. 3.19.

The parameters p and q are defined as

$$p^2 = n_1^2k_o^2 - \beta^2 \tag{3.75a}$$

$$q^2 = \beta^2 - n_2^2k_o^2. \tag{3.75b}$$

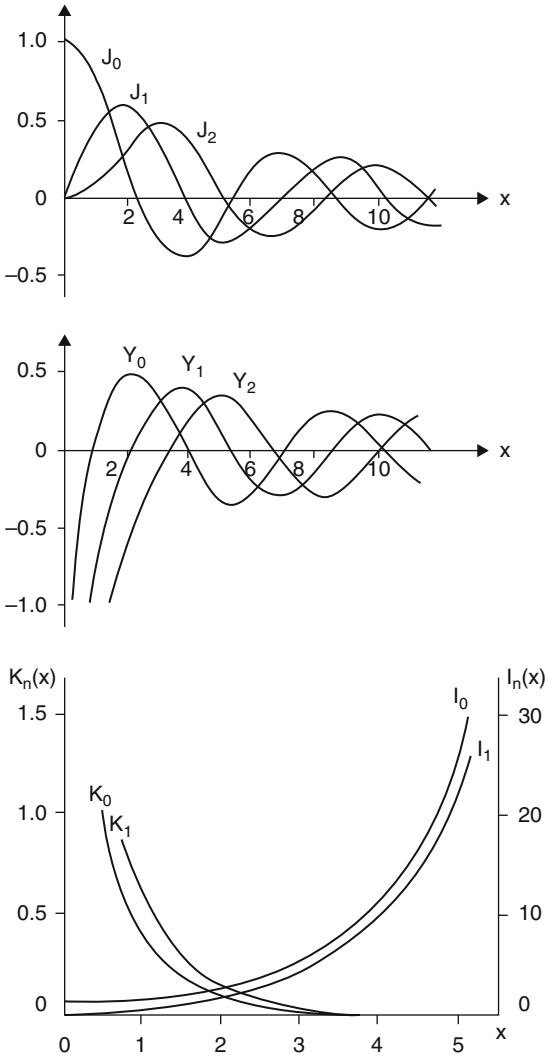


Fig. 3.19 Some Bessel functions

By applying the boundary conditions it results that the optical field for a guided mode should be finite at $\rho = 0$ and go to zero for $\rho \to \infty$. Because Y_m has a singularity at $\rho = 0$, F(0) may stay finite only if $A' = 0$. Similarly $F(\infty) \to 0$ only if $C' = 0$.

The general solution can be written as

$$E_z = \begin{cases} AJ_m \exp(im\phi)\exp(i\beta z) & \rho < a \\ CK_m \exp(im\phi)\exp(i\beta z) & \rho > a \end{cases} \tag{3.76}$$

$$H_z = \begin{cases} BJ_m \exp(im\phi)\exp(i\beta z) & \rho < a \\ DK_m \exp(im\phi)\exp(i\beta z) & \rho > a \end{cases}. \tag{3.77}$$

The other four components may be expressed as

$$E_\rho = (i/p^2)[\beta \partial E_z/\partial\rho + \mu_o(\omega/\rho)\partial H_z/\partial\phi] \tag{3.78a}$$

$$E_\phi = (i/p^2)[(\beta/\rho)\partial E_z/\partial\phi - \mu_o\omega\partial H_z/\partial\rho] \tag{3.78b}$$

$$H_\rho = (i/p^2)\left[\beta\partial H_z/\partial\rho - \varepsilon_o n^2(\omega/\rho)\partial E_z/\partial\phi\right] \tag{3.78c}$$

$$H_\phi = (i/p^2)\left[(\beta/\rho)\partial H_z/\partial\phi + \varepsilon_o n^2 \partial E_z/\partial\rho\right]. \tag{3.78d}$$

The fields in the cladding region are given by the same (3.78) by substituting p with q. In the cladding the field is evanescent. The four constants A,B,C and D are determined by the boundary conditions for continuity of the tangential components of E and H between core and cladding which give four homogeneous equations that have a non trivial solution only if the determinant of the coefficient matrix is zero, that is

$$\begin{aligned}&\left[J_m'(pa)/pJ_m(pa) + K_m'(qa)/qK_m(qa)\right]\left[J_m'(pa)/pJ_m(pa)\right. \\ &\left. + n_2^2 K_m'(qa)/n_1^2 qK_m(qa)\right] = (m^2/a^2)\left[1/p^2 + 1/q^2\right]\left[1/p^2 + n_2^2/n_1^2 q^2\right].\end{aligned} \tag{3.79}$$

For a given set of the k_o,a,n_1 and n_2 parameters this equation may be numerically solved to determine the propagation constant β. Usually multiple solutions are obtained for each m value, β_{mn} each one corresponding to a possible propagation mode. In general both E_z and H_z are not zero (excepted the case m = 0). The effective index is

$$n = \beta/ko. \tag{3.80}$$

Each mode in the fiber propagates with an effective index n that is $n_2 < n < n_1$.

Figure 3.20 shows some modes

The modes HE_{lm} and EH_{lm} are *ibrid* modes, meaning that they have nonzero values of both E_z and H_z. They are defined HE or EH according to the weight of the relative contribution of E_z and H_z to a transverse component of the field in a given point: if E_z gives the larger contribution the mode is EH, if it is H_z to give the larger contribution the mode is HE.

The modes of an optical fibre are sensitive to all perturbations that the evanescent part of the field undergoes.

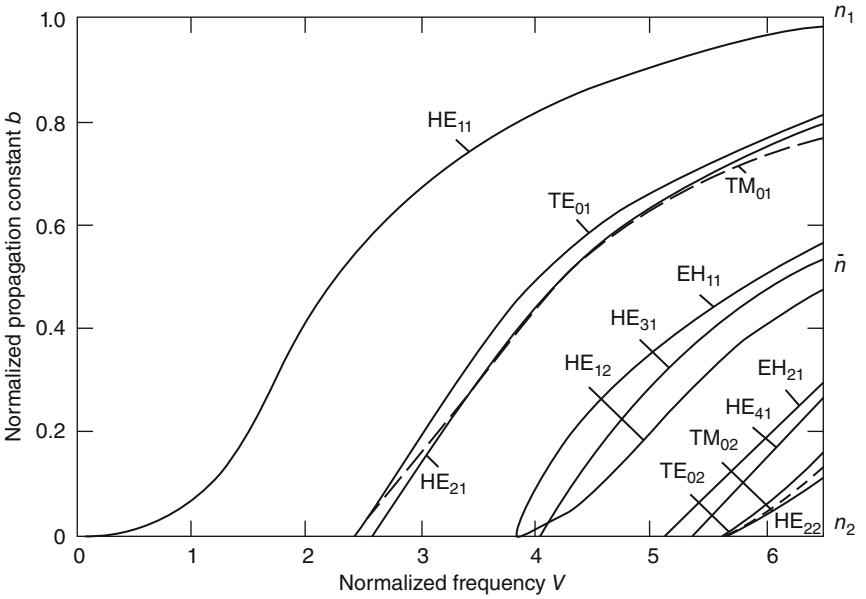


Fig. 3.20 Modes of a step fibre. $V = ka(n_1^2 - n_2^2)½$; $b = \beta/k$

3.6 Multilayers and PBG

Wave propagation along dielectric slab waveguide consisting of multiple parallel layers of dielectrics with different dielectric constants as shown in Fig. 3.21 will now be considered, using a nice formalism introduced by Yeh.[10] This many-layered structure may correspond to a variety of practical structures and may also approximate the case of an inhomogeneous slab with spatially varying refractive index, which can be considered to be formed by many layers of constant refractive index.

Only TE or TM modes may be supported by this structure and we may assume that the expressions for the field components of all modes are multiplied by the propagating factor $\exp i(\omega t - \beta z)$. Dividing the layered structure into m+1 regions as shown in Fig. 3.21 the expressions for the TM fields tangential to the interface in these regions, applying Maxwell equations, are written as

$$\begin{vmatrix} E_z^{(m)} \\ H_y^{(m)} \end{vmatrix} = M_{(m)} \begin{vmatrix} A_m \\ B_m \end{vmatrix} \tag{3.81}$$

[10]Yeh C, Shimabukuro F (2008) The essence of dielectric waveguides. Springer; Yeh C, Lindgren G (1977) Appl Opt. 16:483; Yeh P, Yariv A, Hong CS (1977) J Opt Soc America 67:423

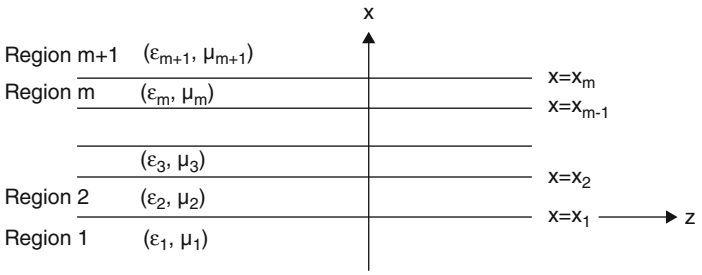


Fig. 3.21 A multilayer stack

where m = 1,2,…m+1, and the matrices M are the following

$$M_{(1)}(x) = \begin{pmatrix} \exp(\beta^2 - k_1^2)^{1/2}x & 0 \\ 0 & -\omega n_1^2\left[1/\exp(\beta^2 - k_1^2)^{1/2}\right]\exp(\beta^2 - k_1^2)^{1/2}x \end{pmatrix} \quad (3.82)$$

$$M_{(m)}(x) =$$

$$\begin{pmatrix} \sin(k_m^2 - \beta^2)^{1/2}x & \cos(k_m^2 - \beta^2)^{1/2}x \\ -i\omega n_m^2\left[1/(k_m^2 - \beta^2)^{1/2}\right]\cos(k_m^2 - \beta^2)^{1/2}x & -i\omega n_m^2\left[1/(k_m^2 - \beta^2)^{1/2}\right]\cos(k_m^2 - \beta^2)^{1/2}x \end{pmatrix} \quad (3.83)$$

$$\mathrm{M_{(m+1)}(x)} = \begin{vmatrix} \exp-(\beta^2 - \mathrm{k_{m+1}^2})^{1/2}\mathrm{x} & 0 \\ 0 & \omega \mathrm{n_{m+1}^2}(1/(\beta^2 - \mathrm{k_{m+1}^2})^{1/2}\exp-(\beta^2 - \mathrm{k_{m+1}^2})^{1/2}\mathrm{x} \end{vmatrix} \quad (3.84)$$

and A_1,B_1,A_2,B_2, … A_{m+1}, B_{m+1} are arbitrary constants. Matching the tangential electric and magnetic fields at the boundary surfaces, that is at x = x_1,x_2,…x_m, and expressing $\begin{bmatrix} A_1 \\ B_1 \end{bmatrix}$ in terms of $\begin{bmatrix} A_{m+1} \\ B_{m+1} \end{bmatrix}$ results in

$$\mathrm{M_1(x_1)}\begin{vmatrix} \mathrm{A_1} \\ \mathrm{B_1} \end{vmatrix} = \mathrm{M_2(x_1)M_2^{-1}(x_2)M_3(x_2)M_3^{-1}(x_3)\ldots M_m^{-1}(x_m)M_{m+1}(x_m)}\begin{vmatrix} \mathrm{A_{m+1}} \\ \mathrm{B_{m+1}} \end{vmatrix} \quad (3.85)$$

where M^{-1} is the inverse matrix[11] of M. (3.85) can be written as

[11]It may help to remember that the elements of M^{-1} are connected to the ones of M by $M_{11}' = M_{22}/\Delta$, $M_{21}' = -M_{21}/\Delta$, $M_{12}' = -M_{12}/\Delta$, $M_{22}' = M_{11}/\Delta$ where Δ is the determinant of the matrix M

$$\begin{pmatrix} \exp(\beta^2-k_1^2)^{1/2}x & -(N_{11}+N_{12}) \\ -\omega n_1^2\left[1/\exp(\beta^2-k_1^2)^{1/2}\right]\exp(\beta^2-k_1^2)^{1/2}x & -(N_{21}+N_{22}) \end{pmatrix} \begin{bmatrix} A_1 \\ B_{m+1} \end{bmatrix} = 0 \quad (3.86)$$

where

$$N = \begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix} = M_2(x_1)M_2^{-1}(x_2)\ldots M_m^{-1}(x_m)M_{m+1}(x_m). \quad (3.87)$$

Setting the determinant of (3.86) to zero, we obtain the dispersion relation from which the propagation constants of various modes along a layered slab may be found.

The TE modes can be treated in a similar way.

We may also consider a wave that crosses a periodic dielectric structure made alternating low and high refractive index layers with a periodicity length comparable to the wavelength of the radiation (Fig. 3.22). This is typically the structure of interference filters, well known and studied since more than one century.[12]

While studying interference filters, the attention was mainly given to their reflection and transmission properties, with no attention to what happens to the field inside the structure. A close examination shows that the periodicity of the structure affects the properties of photons in much the same way that ordinary semiconductor and conductor crystals affect the properties of electrons.

The new name for structures in which scattering or diffracting elements are arranged in such a way that their mutual distances are comparable with the

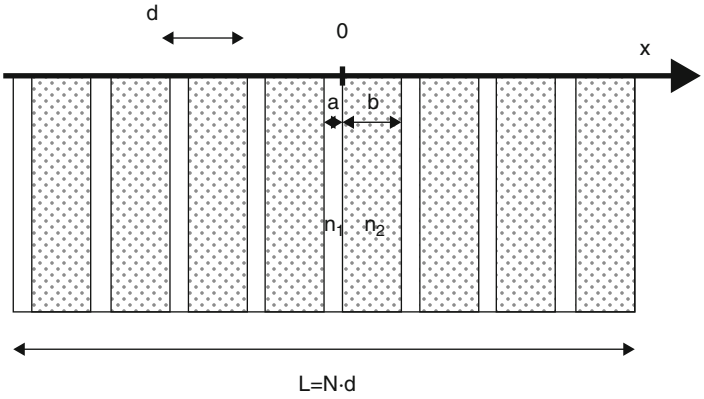


Fig. 3.22 A sketch of a one dimensional layered structure

[12]Macleod HA (2010) Thin-film optical filters. CRC Press

wavelength of the interacting wave, is *photonic band gaps* (PBG's).[13] They have attracted much attention, because they offer exciting ways of manipulating photons allowing control of the propagation of light. It may happen that radiation is trapped inside and some wavelengths are not allowed to pass, so that one may have for the wave-vector a situation similar to the one found for the electron wave-vector of the de Broglie waves in a crystal. It is mainly from this analogy that the name of photonic band gap comes.

The effects are the result of interference between the waves that propagate in one direction and the ones propagating in the opposite direction. To allow these interferences to occur it is necessary that the incoming light be a continuous wave or, if pulsed, the pulse length be much longer than the length of the structure.

The mathematical treatment of 1,2 or 3D structures has different complexity, although the general method is essentially the same.[14] In the case of 2 and 3D structures the name *photonic crystals* is most used.[15] For the sake of simplicity we will here treat only the one dimensional case considering first a periodic structure of infinite extent, and then the more realistic case of a finite structure. In infinite periodic structures the monochromatic field is assumed to be a continuous wave of infinite extent.

3.6.1 Infinite Periodic Structure

In periodic structures the refractive index changes periodically in space, so that the electromagnetic field solution should be in the form of a Bloch wave[16] as is in the case of solid state theory. It means that the field at any plane normal to the propagation direction z along the propagating path is identical to the field at another normal plane, which is a distance, d, away except for a propagation factor $e^{-i\beta d}$. Here, d is the periodicity length of the periodic medium and β is the propagation constant. In other words

$$E(x, y, z + d) = e^{-i\beta d} E(x, y, z) \tag{3.88a}$$

$$H(x, y, z + d) = e^{-i\beta d} H(x, y, z). \tag{3.88b}$$

This is a statement of the Floquet's theorem.[17]

[13] Joannopoulos JD, Villeneuve PR, Fan S (1997) Nature 386:143; John S (1987) Phys Rev Lett. 58:2486; Yablonovich E, Gmitter TJ (1991) Phys Rev Lett. 67:2295

[14] See for ex. Sakoda K (2005) Optical properties of photonic crystals, 2nd edn. Springer

[15] Yablonovitch E (2007) Optics and photonic news, p 12

[16] See Brillouin L (1953) Wave propagation in periodic structures, 2nd edn. Dover

[17] Floquet (1883) Ann de l'Ecole Norm Sup. (2) XII:47; Whittaker ET, Watson GN (1952) A course in modern analysis. Cambridge; Brillouin L (1953) Wave propagation in periodic structures. Dover, New York

The general Helmoltz equation which must be solved is

$$d^2\Psi/dx^2 + (\omega^2/c^2)n^2(x)\Psi = 0. \tag{3.89}$$

Equation (3.89) is Mathieu's equation if $n^2(x)$ is a periodic function which contains only a cosine term

$$n^2(x) = C_o + 2C_1\cos(2\pi x/d) \tag{3.90}$$

where d is the period and C_o and C_1 are suitable constants.

With a change of variable

$$\xi = \pi x/d. \tag{3.91}$$

Equations (3.89) with (3.90) may be written as

$$d^2u/d\xi^2 + (\eta + \gamma\cos 2\xi)u = 0 \tag{3.92}$$

with

$$\eta = (\omega d/c\pi)^2 C_o \quad ; \quad \gamma = 2(\omega d/c\pi)^2 C_1. \tag{3.93}$$

Floquet discovered that the generic solution of the equation could be written as

$$u = D_1 A(\xi)e^{\mu\xi} + D_2 B(\xi)e^{-\mu\xi} \tag{3.94}$$

with amplitudes A and B that are periodic functions of ξ with period π. This solution is the superposition of two waves propagated in opposite directions if μ is imaginary, as can be seen adding the time term. D_1 and D_2 are arbitrary constants.

Considering Fig. 3.22, in the region $-a < x < 0$, $n = n_1$, so

$$(\eta + \gamma\cos 2\xi) = (\omega n_1/c)^2 = -\chi_1^2 \tag{3.95}$$

and in region $0 < x < b$, $n = n_2$, so that

$$(\eta + \gamma\cos 2\xi) = (\omega n_2/c)^2 = -\chi_2^2. \tag{3.96}$$

The solutions of (3.92) are therefore

In the first region $-a < x < 0$

$$u = A\exp(\chi_{1x}) + B\exp(-\chi_{1x}). \tag{3.97a}$$

In the second region $0 < x < b$

$$u = C\exp(\chi_{2x}) + D\exp(-\chi_{2x}). \tag{3.97b}$$

The entire solution must fit the form given by Floquet's theorem. Choosing one of the Floquet's exponential, we write

$$u(x) = A(x)e^{\mu x} \tag{3.98}$$

where A(x) has period d. This means for instance that

$$u(x-d) = A(x-d)e^{\mu(x-d)} = A(x)e^{\mu x}\ \exp^{-\mu d} = u(x)\exp^{-\mu d} \tag{3.99}$$

or

$$u(x) = u(x-d)e^{\mu d}. \tag{3.100}$$

Using this relation, we write the solution in the first interval

$$0 < x < b \quad u(x) = C\exp(\chi_2 x)e^{\mu d} + D\exp(-\chi_2 x)e^{\mu d} \tag{3.101}$$

and the solution in the second interval

$$b < x < b + a = d \quad u(x) = A\exp[\chi_1(x-d)]e^{\mu d} + B\exp[-\chi_1(x-d)]e^{\mu d}. \tag{3.102}$$

The coefficients A,B,C and D can be found using the continuity conditions for u and du/dx at

$$x = 0 \quad A + B = C + D \tag{3.103a}$$

$$A\chi_1 - B\chi_1 = C\chi_2 - D\chi_2 \tag{3.103b}$$

and

$$x = b \quad A\exp(\chi_1 a)e^{\mu d} + B\exp(-\chi_1 a)e^{\mu d} = C\exp(\chi_2 b)e^{\mu d} + D\exp(-\chi_2 b)e^{\mu d} \tag{3.104a}$$

$$A\chi_1\exp(\chi_1 a)e^{\mu d} - B\chi_1\exp(-\chi_1 a)e^{\mu d} = C\chi_2\exp(\chi_2 b)e^{\mu d} + D\chi_2\exp(-\chi_2 b)e^{\mu d}. \tag{3.104b}$$

This is a set of four linear homogeneous equations in A,B,C and D that can be solved if the determinant is zero. After some algebra we find

$$e^{2\mu d} + e^{\mu d}\{\cosh(\chi_1 a)\cosh(\chi_2 b) + (½)[(\chi_1/\chi_2) + (\chi_2/\chi_1)]\sinh(\chi_1 a)\sinh(\chi_2 b)\} + 1 = 0. \tag{3.105}$$

This second order equation has two solutions $e^{\mu d}$ and $e^{-\mu d}$ whose product is one and whose sum is

$$\cos \mu d = \cos \beta = \cosh(\chi_1 a)\cosh(\chi_2 b) + (½)[(\chi_1/\chi_2) + (\chi_2/\chi_1)]\sinh(\chi_1 a)\sinh(\chi_2 b). \tag{3.106}$$

The quantity $\mu d = \beta$ is usually indicated as *Bloch phase*.

According to whether χ_1 and χ_2 are real or imaginary, some of the hyperbolic sines and cosines may become ordinary sines and cosines. We may distinguish three cases:

A. if $\cosh(\mu d) = e^{\mu d} + e^{-\mu d} \geq 1$ it means that μ is a real quantity
B. if $-1 \leq \cosh(\mu d) \leq 1$ it means that μ is purely imaginary
C. if $-1 \leq \cosh(\mu d)$ than $\mu = \mu o + i\pi$.

Case B gives waves propagating through the whole structure without attenuation, which means *passing bands*. Both cases A and C yield waves attenuated exponentially; this means *stopping bands*.

A curve μ vs ω can be obtained from (3.106) which is the dispersion curve of the structure. The transmission band is obtained for the frequencies for which case B is satisfied. In all other cases, being the structure infinite, no radiation may travel in it. In case B, (3.106) can be written with $\mu d = \beta = ikd$, where k is usually indicated as *Bloch wave-vector*

$$\begin{aligned}\cos(kd) = {} & \cos(\omega n_1 a/c)\cos(\omega n_2 b/c) + (½)[(n_1/n_2) \\ & + (n_2/n_1)]\sin(\omega n_1 a/c)\sin(\omega n_2 b/c).\end{aligned} \tag{3.107}$$

Fig. 3.23 shows the dispersion relation near a stop-band. By increasing the frequency, k increases up to the value corresponding to $kd = \pi$.

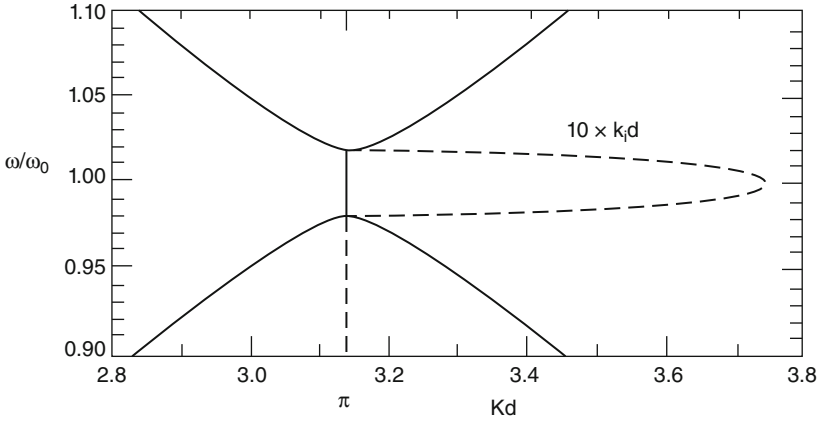


Fig. 3.23 Dispersion relation near a forbidden band

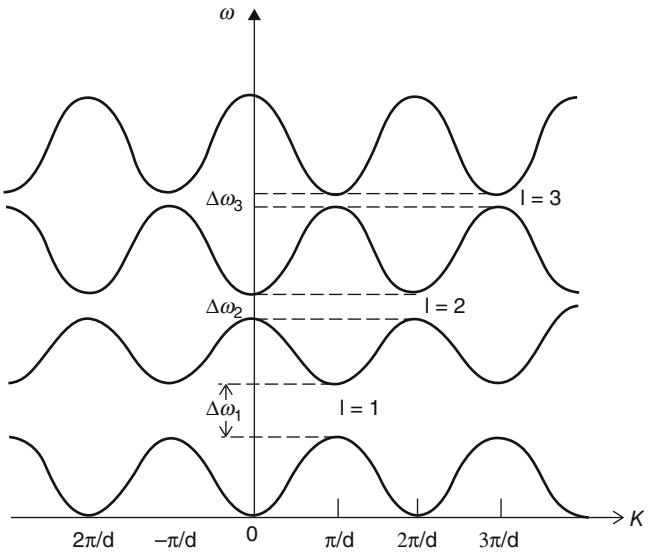


Fig. 3.24 A general dispersion relation showing the fundamental (l = 1) and higher order (l = 2,3) forbidden bands

For still increasing frequencies, k becomes complex. We see that the relation (3.107) repeat itself in the regions $-\pi < kd < \pi$, $2\pi < kd < 3\pi$, and so on. The regions are the so-called *Brillouin zones*.

Therefore it is possible to show the general dispersion relation ω(k) showing the fundamental (l = 1) and higher order (l = 2,3,..) forbidden bands which repeat periodically with period 2π/d (Fig. 3.24). The first Brillouin zone ranges between $k = -\pi/d$ and $k = +\pi/d$.

3.6.2 Finite 1D PBG

The preceding discussion refers to an infinite periodic structure. If we consider only a finite number of layers, we may treat the problem using transfer matrices according for example to the method described in paragraph 3.6 The traditional transfer matrix method applied to this problem is described for example in ref.[18] If we are interested only to the reflection and transmission of the structure we may use a simplified method.[19] The general solution of the Helmoltz (3.89) can be

[18]Bertolotti M (2001) Nanoscale linear and nonlinear optics. In: Bertolotti M, Bowden CM, Sibilia C (eds.) AIP Conf Proc. New York, p 1

[19]Bendickson JM, Dowling JP, Scalora M (1996) Phys Rev. E53:4107

written as a superposition of right [u+(x)] and left [u − (x)] going waves which can be put in the column form

$$u = \begin{vmatrix} u_+(x) \\ u_-(x) \end{vmatrix} \tag{3.108}$$

with the boundary conditions

$$u(0) = \begin{vmatrix} 1 \\ r \end{vmatrix} \quad u(d) = \begin{vmatrix} t \\ 0 \end{vmatrix} \tag{3.109}$$

where r and t are the reflection and transmission coefficients respectively of the cell. A transfer matrix may be defined

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{3.110}$$

defined through

$$u(x) = Mu(x - d) \tag{3.111}$$

which should satisfy the condition

$$u(0) = Mu(d) \tag{3.112}$$

which gives

$$A = 1/t \quad C = r/t. \tag{3.113}$$

To derive the B and D coefficients one may observe that for a complex wave $E = e^{ikx}$ the operation of time reversal τ corresponds to complex conjugation

$$\tau E = E^* \tag{3.114}$$

For real, linear index profiles the process must be invariant under time reversal and therefore the incoming field is r* and the reflected is 1 at boundary x = 0, so that

$$\begin{aligned} u^*(0) &= \begin{bmatrix} r^* \\ 1 \end{bmatrix} \\ u^*(d) &= \begin{bmatrix} 0 \\ t^* \end{bmatrix} \end{aligned} \tag{3.115}$$

These conditions mean B = r*/t* and D = 1/t*. So eventually the matrix M can be written as

$$M = \begin{pmatrix} 1/t & r^*/t^* \\ r/t & 1/t^* \end{pmatrix}. \tag{3.116}$$

This is the matrix representing a single cell. It can be shown that for a structure of the type shown in Fig. 3.22 made with N cells, one may write the matrix as

$$M = \begin{pmatrix} 1/t_N & r^*{}_N/t_N{}^* \\ r_N/t_N & 1/t_N{}^* \end{pmatrix} \tag{3.117}$$

where the transmission and reflection coefficients of the N-cell structure can be expressed as a function of the infinite-period Bloch phase β as

$$1/t_N = (1/t)[\sin(N\beta)/\sin\beta] - [\sin(N-1)\beta]/\sin\beta \tag{3.118}$$

$$r_N/t_N = (r/t)(\sin N\beta/\sin\beta). \tag{3.119}$$

The N period transmittance

$$T_N = |t_N|^2 \tag{3.120}$$

is

$$1/T_N = 1 + [\sin(2N\beta)/\sin(2\beta)][(1/T) - 1] \tag{3.121}$$

where T is the unit cell transmittance.

The infinite-period Bloch phase β determines the transmission of the finite structure. In particular the pass-band and the stop-bands for the N-period mirror very closely those of the infinite lattice. In the pass-band (β real) T_N varies sinusoidally with β and ω. The behaviour of transmission as a function of ω is therefore strongly dependent on the number of periods and tends to become steeper as N increases.

Figure 3.25 shows the general form of the transmission (left) and reflection (right) of a structure made using layers whose optical thickness was such to be λ/4 of the wavelength at which one wishes to have the maximum reflection.

Inside the gap, β is complex and the sinusoidal behaviour changes into a hyperbolic exponential form. The transmission coefficient for a N periods structure is

$$t_N = x_N + iy_N. \tag{3.122}$$

The phase on transmission is

$$\varphi = (y/x)\tan(N\beta)\cot\beta \tag{3.123}$$

where x and y are the real and imaginary part of the single cell transmission.

A useful phenomenological way to treat propagation in a finite structure is to introduce an effective index n_{eff} defined referring to the phase of the wave transmitted

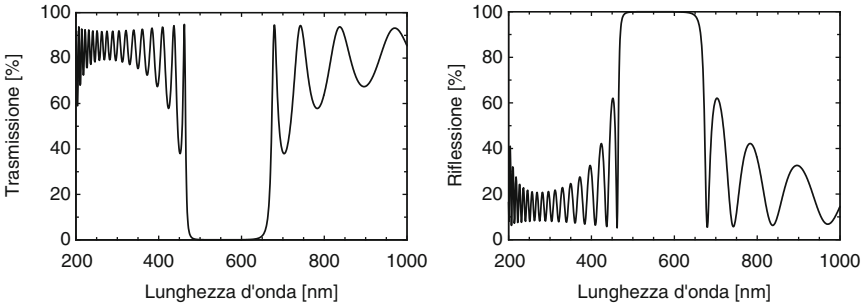


Fig. 3.25 An example of a transmission spectrum (on the left) and a reflection spectrum (on the right)

through the structure. One may consider the phase φ as the one that the wave should suffer traversing an equivalent homogeneous structure of the same length L = Nd

$$\varphi = n_{eff} Lk = n_{eff} L\omega/c = k(\omega)L \tag{3.124}$$

where k(ω) is the effective wavevector. Note that this phase φ is the total phase shift of the structure. If it was distributed uniformly along the N cells, one could consider the shift per cell φo and write down φ = Nφo. One could be tempted to say that φo = β the Bloch phase, but this is wrong because truncation of the structure changes the phase shift suffered by a single cell and moreover makes his phase shift to depend on the cell position along the structure.

One may write for the field of a plane wave traversing the structure

$$E_{transmitted} = A\exp(in_{eff}\omega/c) \tag{3.125}$$

defining

$$n_{eff} = n_{eff} + ic\gamma/\omega \tag{3.126}$$

being γ the attenuation

$$\exp(-2\gamma L) = |t|^2. \tag{3.127}$$

At the end, the effective index can be written

$$n_{eff} = (c/\omega L)\left[tg^{-1}(y_N/x_N) \pm m\pi - (i/2)\ln|t|^2\right] \tag{3.128}$$

with the integer m that can be expressed as a function of Bloch phase as

$$m = \text{int}[(N\beta/\pi) + ½] \tag{3.129}$$

where int is the integer part, with a range of definition of the tangent function between $-\pi/2$ and $+\pi/2$.

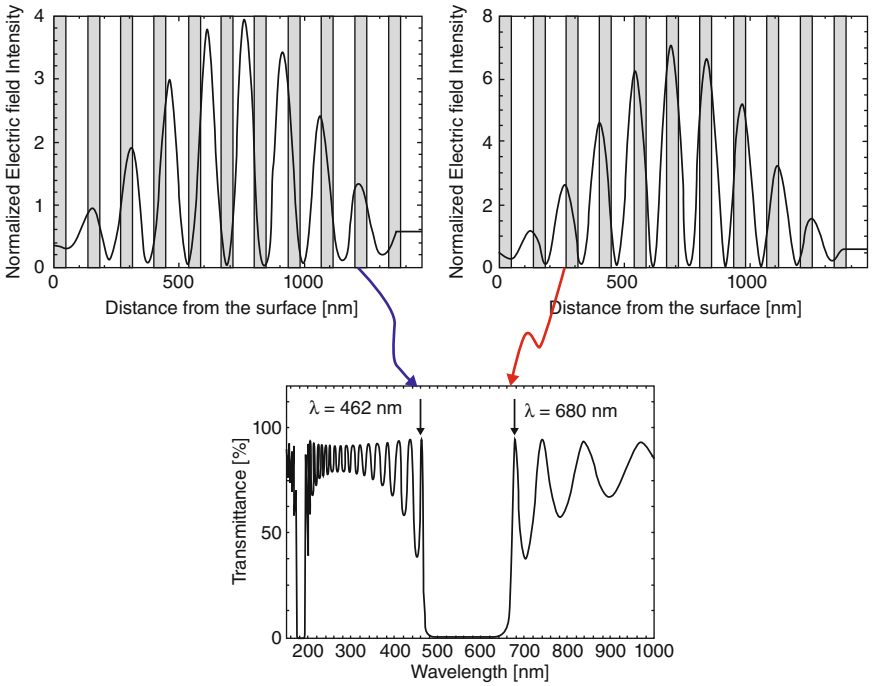


Fig. 3.26 Localization of the modulus of the electric field inside a 1D PBG. Note that the field localizes in the high (grey) or low (White) refractive index regions according to which peak of transmission is selected

The truncation of a finite structure has two main results. The first is that a monochromatic wave due to the finite dimension broadens in k.[20] In an one-dimensional structure, assuming a mean field approach, the structure may be considered as a bulk material with some effective refractive index n_{eff} whose dispersion is governed by the geometry of the structure and, of course, the natural dispersion of the material.

The second, is that the field becomes localized in the structure. This localization increases by increasing the number of cells. This is an effect similar to the enhancement which is present in a Fabry-Perot by increasing the reflectivity of mirrors. Paradoxically, by increasing N the localization of the electronic field increases and the infinite structure situation in which the field is not localized is never reached.

To have localization it is necessary that the space length of the electromagnetic pulse be much longer than the length of the structure, so allowing interference effects to take fully place.

Inside the structure the electric field is distributed in a selective way. At some frequency the maximum electric field is localized in the high refractive index regions and at a different wavelength in the low index region (Fig. 3.26). It is possible to have great localization of the field in regions that may be suitably selected.

[20]Centini M et al (2000) J Opt Pure and Appl Opt. 2:121

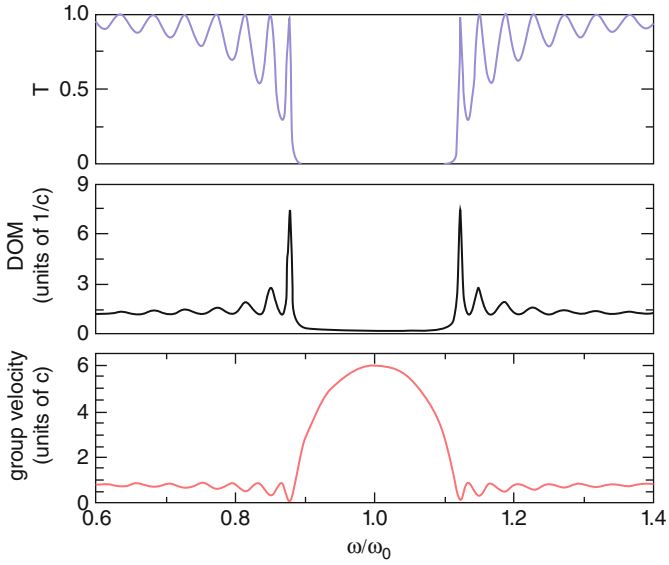


Fig. 3.27 Group velocity and density of modes in a 1D PBG whose transmission spectrum is shown in the upper part of the figure. Note the enhancement of the mode density at the first transmission peaks near the gap with the corresponding minima of group velocity that becomes superluminal at the centre of the band gap

A spectral density of electromagnetic modes inside the structure can be defined as the inverse of the group velocity

$$S(\omega) = dk/d\omega = 1/v_g. \tag{3.130}$$

The density of modes reflects the form of the transmission increasing as the contrast n_2/n_1 increases (Fig. 3.27). As a consequence the group velocity decreases showing a minimum in correspondence of each maximum in the transmission.

In conclusion the main properties of the structure probed with white continuous light are:

- inhibited regions in the transmission spectrum (*band gap*);
- oscillations in the transmission and reflection spectrum whose number is given by the number N of periods of the structure;
- dependence of the band gap from the building criteria. The amplitude of band gap depends on the index contrast $n_H - n_L$ and the steepness of its edges from the number N of periods;
- inside the structure the electric field is distributed in a selective way.

The peculiar properties of these structures allow a number of interesting effects. For example, the existence of a gap in the density of photonic modes gives the possibility to inhibit or enhance spontaneous emission.

Photonic band gaps have also very interesting nonlinear properties we will not discuss here.[21]

3.7 The Role of Evanescent Waves in Waveguide Sensors

Attenuated total reflection is often used to make sensors. The radiation that strikes the interface of two dielectric media from the optically denser medium at an angle of incidence larger than the critical angle is totally reflected back into the optically denser medium but the reflectivity at the interface is a function of the optical properties of the optically rarer medium. Thus, the totally reflected radiation may pick up information from the medium adjacent to the waveguide.

An example of this behaviour is shown in Fig. 3.28 in which a system to register the fingerprints using total reflection at one side of a glass prism is sketched.

The ridges on a finger act as locations where TIR is frustrated, so no light comes from there. But between the ridges, there is still TIR so light is reflected. A digital camera collect the totally reflected light and recover the image.

In sensors, the optical parameter whose change is utilized can be the refractive index (the change of the real part is utilized), the absorption (the imaginary part of the refractive index) or the luminescence (the evanescent field excites luminescent molecules). Luminescent sensors are in general chemical sensors, meaning that they enable measurement of chemical concentration.

If the refractive index of the material has a functional dependence from wavelength, the method allows spectroscopic measurements. If an absorbing medium is placed in contact with the surface, the reflectance for total internal reflection is attenuated at the absorption wavelength of the sample.[22]

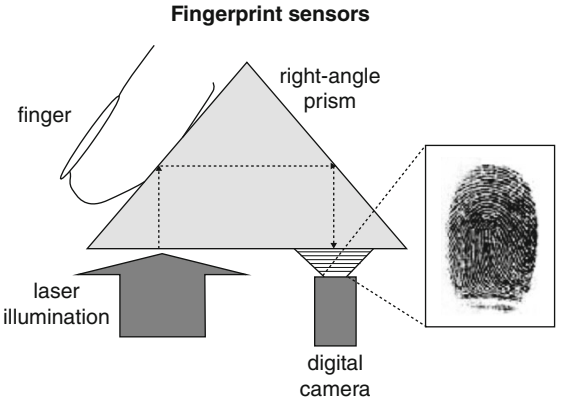


Fig. 3.28 Schematic of a fingerprint sensor

[21] See for example Bertolotti M (2006) J Opt Pure and Appl Opt. 8:S9

[22] Polerecky L et al (2002) Appl Opt. 41:2879

All molecules close to interface absorb energy from the evanescent field at their specific absorption wavelength, and the radiation transmitted through an ATR (attenuated total reflection) element shows an absorption pattern similar to standard absorption spectra derived from transmission experiments. The exponentially decreasing evanescent field especially allows to monitor information quite close to the interface to the waveguide.

Roughly speaking, each attenuated total reflection provides the same information as a transmission cell with an optical path-length equalling the wavelength of the totally reflected photon. Consequently, it is possible to build sensor heads with transmission path-lengths of just a few μm to some hundred μm.

Sensors can be made with planar waveguide structures or fibers.

A range of sensing configurations include passive waveguides which exploit evanescent wave effects when light undergoes total internal refection at the interface between the guiding region and the cladding. A variety of planar waveguide configurations are currently employed. These include thin film (low-mode) waveguides and ridge waveguides.

However, bearing in mind the demand for low cost platforms, the most practical configuration is that of a multimode slab waveguide, which can take the form of a simple plastic or glass slide. Figure 3.29 illustrates a typical planar optical sensor platform based on such waveguide configuration. Light is coupled into the waveguide and propagates along its length where it is then out-coupled to an appropriate detector. For chemical sensing the waveguide is coated with an analyte-sensitive layer that exhibits changes in the optical properties in response to the presence of the analyte. The nature of these changes is dependent on the optical phenomenon employed, e.g. optical absorption or fluorescence effects.

As the mode propagates within the waveguide by total internal reflection, its exponential decaying evanescent tail extends into both cover and substrate layers over a distance that is characterised by the penetration depth d_p which can be expressed as

$$d_p = \lambda_o / 2\pi n(\sin^2\theta - \sin^2\theta_c)½ \tag{3.131}$$

where n is the refractive index of the waveguide and θ is the angle of incidence of the propagating light at the waveguide-cover layer interface.

Figure 3.29 illustrates a sensor which consists of a planar waveguide on top of which a thin sensing layer has been deposited. Light propagates along the waveguide by total internal reflections with an incidence angle greater than the critical angle. The evanescent field of the propagating light interrogates the sensing layer and the output intensity can be written as

$$I = I_o \exp(-rCL_{int}) \tag{3.132}$$

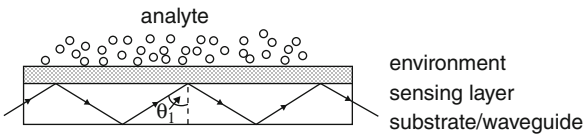


Fig. 3.29 A generalized planar optical waveguide-based sensor

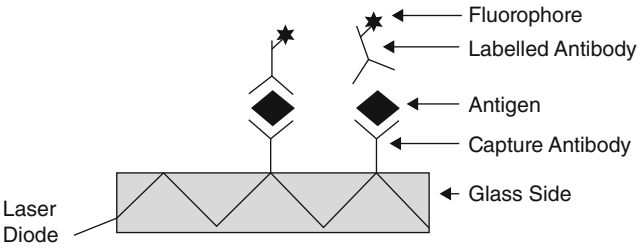


Fig. 3.30 A planar platform for evanescent-wave excitation of fluorescence

where C is a constant, L_{int} is the interaction length of the waveguide and r is the ratio of the optical power present in the evanescent field to the total guided optical power. In this configuration, sensitivity is proportional to the product rL_{int}. Consequently, improved sensitivity can be achieved with an evanescent wave-based platform when compared with a configuration based on transmittance the product rL_{int} is greater than the optical path length (i.e. sensing layer thickness) L, in the latter configuration. It is important to note that this improved performance can be achieved while still employing a thin sensing layer, meaning that sensor response time is not adversely affected.

A more complex system is shown in Fig. 3.30 showing a biosensor platform which consists of a multimode slab waveguide on the upper surface of which antibodies have been immobilised. These antibodies selectively bind the target analyte, which, in this case is a specific antigen present in the sample solution that is passed over the platform using the required fluidics. A second solution containing fluorescently labelled antibodies similar to those bound to the surface of the waveguide is subsequently passed over the platform. These antibodies again bind selectively to the surface-bond antigens. Light from a suitable source is coupled to the glass side and propagates along its length by total internal reflection. The evanescent field of the light extends above the surface of the waveguide where it excites fluorescence from the bound, labelled antibodies. The fluorescence from the upper surface is imaged using an appropriate detector (typically a CCD camera) and the recorded fluorescence is correlated to antigen concentration. The tuneable nature of the evanescent field penetration depth is critical to the effective operation of this sensor as it facilitates surface-specific excitation of fluorescence. This means that only those fluorophores attached to the surface via the antibody-antigen-labelled antibody recognition event (i.e. those within the distance defined by the penetration depth of the evanescent field) are excited, permitting discrimination between surface binding events and fluorescence from the bulk solution.

Another possibility to interrogate the effect of the evanescent field on the propagation of the guided radiation in the waveguide is to use a Mach-Zehnder interferometer[23] (Fig. 3.31).

[23]Brandenburg A, Henninger R (1994) Appl Opt. 33:5941

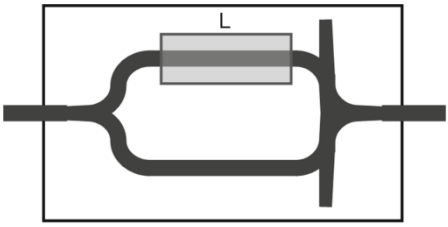


Fig. 3.31 An interferometric sensor. The incoming light intensity Io is connected to the output intensity I(ΔΦ) by $I(\Delta\Phi)/Io = (1/2)[1 + \cos(Lk\Delta n)]$

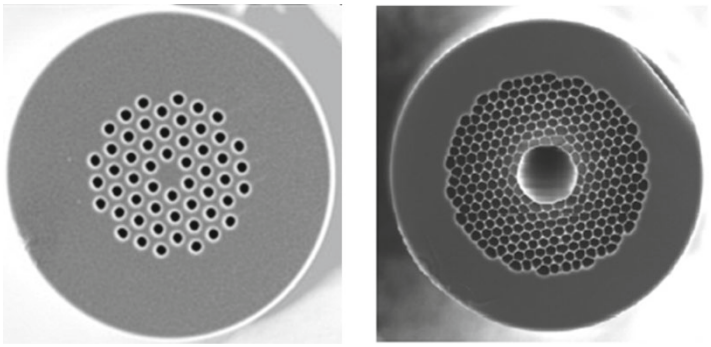


Fig. 3.32 Two different geometries of the microstructured fibres. On the left the geometry with a central solid core and on the right the central hollow type

The wave guide is split into two arms; one arm (the reference arm) is shielded and not affected by the environment. At the interface of the other arm, the evanescent field will be influenced by any sensing effect. The two beams will propagate in a different way within the two arms, in dependence on the difference in the refractive indices adjacent to the arms. This causes a phase difference ΔΦ which results in a change in the position of the fringes at the end of the interferometer that can be detected.

Fibre Evanescent Wave Spectroscopy (FEWS) is based on the same principle of conventional ATR, but uses an optical fibre as a waveguide. Fibres allow to increase the number of total internal reflections, which in turn increases the sensitivity of the sensor. To allow the sensing, the fibre should be stripped of the cladding

Entirely fibre-optic systems are generally more flexible and, both optically and mechanically, easier to design and align than systems involving rigid optics. Finally, fibre-optic sensors are significantly cheaper.

A vast literature exists on evanescent field sensors.[24]

A new class of optical fibre emerged in 1996 under the name of *holey fibres* or *microstructured fibres*. In the cladding of a microstructured fiber, there is a periodic alignment of holes that are approximately 1 μm in diameter. These are grouped around either a solid glass core or a large, hollow space (Fig. 3.32). The

[24]Boisde G, Harmer A (1996) Chemical and biochemical sensing with optical fibers and waveguides. Artech House, Boston; Ligler FS, Rowe Tait CA (2002) Optical biosensors: present and future. Elsevier, Amsterdam; see also Sirbuly DJ et al (2007) Adv Mater. 19:61

two structures represent two types of microstructured fiber, which differ in principle and in application. In the first case, the core of these fibres is formed by a solid silica region, and the cladding is comprised of air holes which run along the fibre.[25] Light guiding was initially assumed to occur because light was guided using photonic band effects.[26] It was soon realised that total internal reflection was the leading mechanism because the index of the core was higher than the effective index of the cladding. In the second type of fiber,[27] the hollow core one, light is guided via photonic band gap effects.[28]

Holey fibers, in which air holes were introduced in the cladding region were considered since the early days of silica waveguide research.[29] They were successively not developed and the idea was resumed nearly twenty years later.

With careful choice of fiber parameters, such as lattice pitch and air-hole diameter, photonic crystal fibers (PCF)[30] in which a pattern of air holes is running along the entire length of the fiber have a large fraction of the optical field propagating through the fiber as an evanescent field.[31] It is therefore possible put the material under investigation into the air holes. These fibres can be used for evanescent-wave sensing of gases[32] or biomolecules.[33] In conventional fibres one must remove the coating and cladding to obtain an overlap between the optical field and the sample. The air holes of the PCF can instead be filled with the measuring material which finds itself in regions with a strong optical field, while the fiber can keep both the cladding and the coating on, thereby ensuring a robust device.

A full description with the many variants of these fibres was made by Russell.[34]

[25]Birks TA et al (1995) Electron Lett. 31:1941; Birks TA, Knoght JC, Russell P St. J (1997) Opt Lett. 22:961

[26]Birks TA et al (1995) Electron Lett. 31:1941

[27]Knight JC et al (1998) Science 282:1475; Cregan RF et al (1999) Science 285:1537

[28]Birks TA et al (1995) Electron Lett. 31:1941; Cregan RF et al (1999) Science 285:1537

[29]Kaiser PV, Astle HW (1974) Bell Syst Tech J. 53:1021

[30]Knight JC et al (1996) Opt Lett. 21:1547

[31]Monro TM et al (1999) Electron Lett. 35:1188

[32]Hoo YL et al (2003) Appl Opt. 42:3509

[33]Jensen JB et al (2004) Opt Lett. 29:1974

[34]Russell P St. J (2006) J Lightwave Technol. 24:4729

Chapter 4
High Resolution Optical Microscopes

4.1 Introduction

Several methods which use in one way or the other evanescent waves have been devised to overcome the classic resolution limit established by Abbe[1] and Lord Rutherford[2] as we saw in Chap. 2, according to which the best resolution achievable with optical light is about 200 nm.

By taking into account knowledge about the excitation properties or the spectral properties of the sample it is possible to stretch the resolution limit beyond the diffraction limit.[3] Also a number of nonlinear processes like two photon absorption,[4] second harmonic generation or CARS (Coherent Anti-Stokes Raman Scattering), have been devised which allow very high resolution. Resolution can be extended beyond Abbe's limit also by driving the light matter interaction into saturation[5] or by recording higher harmonics of the sample's spatial frequencies.[6] Stimulated emission depletion (STED) quenches the fluorescence surrounding a very small volume, thus, effectively increasing the resolution in both lateral and axial directions to about 100 nm. With this technique Hell and co-workers have achieved an optical resolution of 33 nm.[7]

Here we limit the discussion to those systems which use evanescent waves. These are generically called *near-field microscopes*. In this class are SNOM (Scanning Near-field Optical Microscopy), STOM (Scanning Tunnelling Optical

[1]Abbe E (1873) Archiv f Mikroskop Anat. 9:413

[2]Lord Rayleigh (1896) Phil Mag. 5(42):167

[3]Toraldo di Francia G (1955) J Opt Soc America 45:497

[4]Denk W et al (1990) Science 248:73

[5]Klar TA et al (2000) Proc Nat Acad Sci U S A. 97:8206; Willig KI et al (2006) Nature 440:935

[6]Gustafsson MGL (2005) Proc Nat Acad Sci U S A. 102:13081

[7]Klar TA et al (2000) Proc Nat Acad Sci U S A. 97:8206; Klar TA et al (2001) Appl Phys Lett. 78:393; Dyba M, Hell SW (2002) Phys Rev Lett. 88:163901

© Springer International Publishing AG 2017

M. Bertolotti et al., *Evanescent Waves in Optics*, Springer Series in Optical Sciences,
DOI 10.1007/978-3-319-61261-4_4

Microscopy), and TIRF (Total Internal Reflection Fluorescence Microscope). A short historical discussion on their origin is given by Novotny.[8] An ample literature exist on the subject.[9]

4.2 Scanning Near-field Optical Microscopy (SNOM)

The *Scanning Near-field Optical Microscope* (SNOM) is a purely optical microscopy technique which allows to form images with resolution well beyond the diffraction limit; it combines the excellent properties of classical optical microscopy with a lateral resolution reaching well into the sub 100 nm regime.

To clarify the difference with ordinary microscopes, in a classical far-field optical microscope the object is illuminated by a monochromatic plane wave and the transmitted or reflected light, scattered by it, is collected by a lens and imaged onto a detector. For practical reasons, the lens is placed at least several wavelengths off the object surface, i.e. in the *far field* (Fig. 4.1 left). Since high spatial frequencies corresponding to fine details of the object generate Fourier components of the light field that decay exponentially away from the object (*evanescent waves*), they cannot be collected by the lens. This effect leads to the well known Abbe diffraction limit.

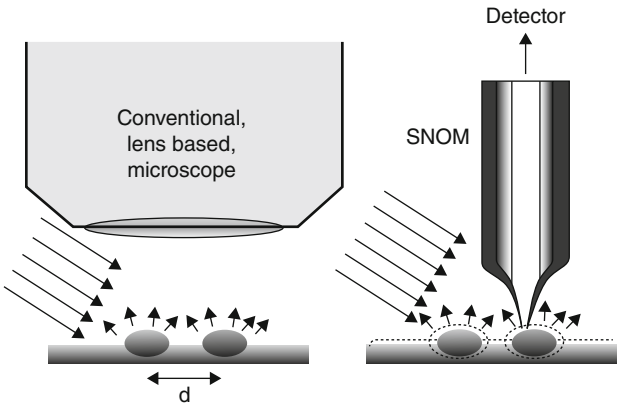


Fig. 4.1 The principle of an "ordinary" microscope (left), and a SNOM (right)

[8]Novotny L (2007) Progr Opt. 50:137

[9]Paesler MA, Moyer PJ (1996) Near field optics: theory, instrumentation and applications. Wiley; Fillard JP (1996) Near-field optics and nanoscopy. World Scientific Publ. Co.; Otshu M, Hori M (1999) Near-field nanooptics. Kluwer; Richard D, Zayat A (2004) Phil Trans R Soc London A362:699–919, special issue; Fisher, Pohl (1989) Phys Rev Lett. 62:458; Novotny, Stranick (2006) Ann Rev Phys Chem. 57:303

With the introduction of the Near-field Scanning Optical Microscope (NSOM) usually called SNOM (Scanning Near-field Optical Microscope) (Fig. 4.1 right), this limitation no longer exists because evanescent waves are caught, and optical resolution of <50 nm can be achieved.

In 1928, E H Synge[10] proposed an experimental scheme that would allow optical resolution to extend into the nanometre regime, using a strong light source behind a thin, opaque metal film with a 100 nm diameter hole in it as a very small light source. The tiny spot of light created in this way, should be used to locally illuminate a thin biological section. In order to guarantee the local illumination, he imposed the condition that the aperture in the metal film be no further away from the section than the aperture diameter, i.e., less than 100 nm. Images were to be recorded point by point raster-scanning the image, detecting the light transmitted by the biological section by means of a sensitive photodetector.

For the small opening, he suggested using either a pinhole in a metal plate or a quartz cone coated with a metal except for the tip (Fig. 4.1 right). He exchanged letters with Einstein who supported him.[11]

The mathematician J A O'Keefe,[12] not knowing the Synge's work, proposed to employ an irradiated small hole in an aluminium coating as a light source and maintained that by scanning, an image can be built up whose details will correspond to the size of the hole, even if it is smaller than the wavelength of the light used.

The research continued.[13] Ash and Nicholls[14] in 1972 gave an experimental demonstration using a 1.5 mm aperture illuminated with 3 cm waves obtaining a resolution $\lambda/60$, while the Rayleigh limit is of the order of $\lambda/2$.

In 1984, shortly after the invention of the scanning tunnelling microscope,[15] nanometre-scale positioning technology was available and an optical microscope similar to Synge's proposed and forgotten scheme was re-invented by D Pohl and demonstrated together with W Denk and U Duering at the IBM Ruschlikon Research Laboratory.[16] Pohl and collaborators obtained an optical image with a lateral resolution of 25 nm at a wavelength of 488 nm, that is a resolution of the order of $\lambda/20$.

Independently, a similar scheme was proposed and developed by Lewis and his group at Cornell University.[17] The key innovation was the fabrication of a

[10]Synge EH (1932) Phil Mag. 6:356; Synge EH (1932) Phil Mag. 13:297

[11]Novotny L (2007) Progr Opt. 50:137

[12]O'Keefe JA (1956) J Opt Soc America 46:359

[13]Baez AV (1956) J Opt Soc America 46:901; McCutchen CW (1967) J Opt Soc America 57:1190; Massey GA (1984) Appl Opt. 23:658

[14]Ash EA, Nichols G (1972) Nature 237:510

[15]Binning G et al (1982) Helv Phys Acta 55:726

[16]Pohl D, Denk W, Lanz M (1986) Appl Phys Lett. 44:651; Duerig U, Pohl D, Rohner F (1986) J Appl Phys. 59:3318

[17]Lewis A et al (1984) Ultramicroscopy 13:227; Harootunian A et al (1986) Appl Phys Lett. 49:674

sub-wavelength optical aperture at the apex of a sharply pointed transparent probe tip that was coated with a metal. Since then several different types of near-field optical probes have been developed.[18]

The central idea of near-field optical microscopy is to use the spatial frequencies associated with evanescent waves with $k_{parallel} > k$. thereby increasing the bandwidth of spatial frequencies. Because evanescent waves attenuate strongly with distance, the sample must be located very near to the source, which means that the sample becomes part of the system and the interaction between source and sample complicates data analysis considerably. Here we do not discuss this coupling mechanism.

To expand the spatial frequency spectrum, an illuminated very small aperture in a metal screen, or a metal tip or a small particle favourably illuminated can be used. In the case of a small aperture, beyond the aperture the field spreads out very rapidly as a consequence of diffraction. This can be seen very easily considering the angular spectrum representation. Assume the x component of the field has a Gaussian amplitude distribution

$$E_x(x, y, 0) = E_o \exp\left[-(x^2 + y^2)/w_o^2\right]. \tag{4.1}$$

The Fourier spectrum is

$$E_x(k_x, k_y, 0) = E_o \exp[-(k_x^2 + k_y^2)w_o^2/4], \tag{4.2}$$

which shows that the smaller is the beam waist w_o, the greater is $k_{parallel} = (k_x^2 + k_y^2)^{1/2}$.

The evanescent field is scattered by the fine structures of the sample being converted in a propagating field which is so detected.

These considerations are very rough and can be used only to give an intuitive insight of the general principle of operation. Each near-field source (tip, aperture, particle, etc) has its own unique field distribution and the unavoidable interaction between source and sample is also different for each source. Elaborate calculations are needed to model all this.

The standard set-up of a SNOM is shown in Fig. 4.2. A point light source is brought at nanometric distance from the surface that will be imaged and is then scanned over the surface, without touching it, while the optical signal from the surface is collected and detected through a small aperture at the tip of a tapered optical fibre that is scanned over the sample (Fig. 4.1 right). The signal is collected from a very small portion of the sample, approximately the same size as the aperture at the optical fibre tip.

Since we have two separate elements for illumination and detection, the two elements must finally share the same optical axis.

[18]Betzig E, Isaacson M, Lewis A (1987) Appl Phys Lett. 51:2088; Kopelman R et al (1991) J Luminesc. 48–49:871; Kuck N et al (1992) Appl Phys Lett. 61:139; Betzig E et al (1992) Appl Opt. 31:4563

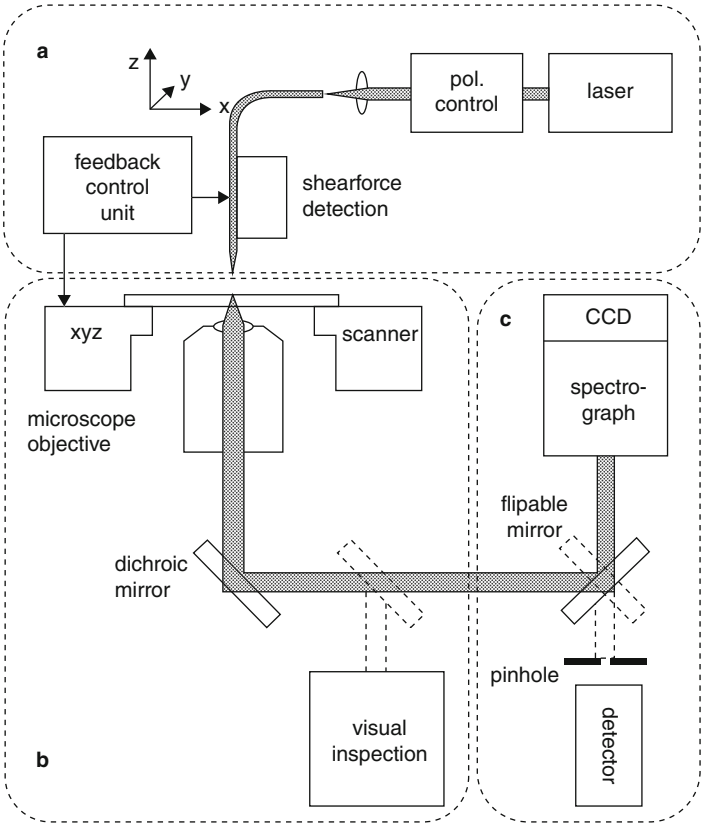


Fig. 4.2 Standard SNOM set-up consisting of (**a**) an illumination unit, (**b**) collection and redistribution unit and (**c**) a detection module. (From B Hecht et al., J Chem Phys. 112, 7761 (2000)[21] reproduced with permission from AIP)

The working operation just described is not unique. There are several principles of operation of SNOM. A first classification can be made according to the mode of collection of the light or the mode of illumination and the microscopes can be indicated as "*aperture*" or "*aperture-less*" SNOM.

In the *aperture mode*, the aperture can either act as a miniature light source when it is illuminated from the back side or it can be a small probe, collecting light from the optical near field in the immediate proximity of the sample.

In the *aperture-less* microscopes a sharp metallic tip scans the sample and scatters propagating fields.[19] In this type of operation the sample is illuminated with a normal propagating wave and a local probe put in the near-field of the sample, that is at a distance smaller then around 100 nm, collects part of the evanescent

[19]Akhremitchev BB et al (2001) Langmuir 17:2774; Akhremitchev BB et al (2002) 18:5325

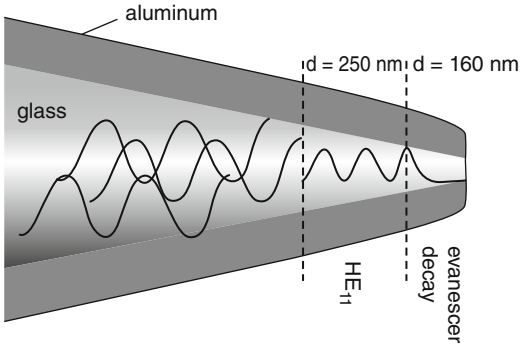


Fig. 4.3 Mode propagation in a tapered metal-coated optical fiber at a wavelength of 488 nm cut-off diameter. (From B Hecht et al., J Chem Phys. 112, 7761 (2000)[21] reproduced with permission from AIP)

field scattered by the sample (Fig. 4.1 right). The role of the local probe (the tip) is to establish a local interaction and to scatter away the local field.

The optical design of an aperture-type scanning near-field optical microscope, that was the one proposed by Singe, uses an evanescent wave to probe the sample. This wave is obtained from a local probe metalized with a nanometric opening at its end (Fig. 4.3). The light propagation in such structures is very complex.[20]

The radiative wave propagating in the probe becomes evanescent when it passes through the nanometric opening. If the sample is put in the near-field of the probe, the sub-wavelength roughness of the sample converts part of the evanescent field back into a propagating wave that can be collected in the far-field. The principal difficulty is to collect the radiative waves because their intensity is very low.

Most commonly, apertures are formed by coating the sides of a sharply pointed optical fibre with metal. The uncoated apex of the pointed fibre represents an aperture. Its size is defined by the sharpness of the fibre tip and the quality of the metal coating (Fig. 4.4). The optical response due to the interaction between the near-field and the sample surface is recorded by a lens.

It is also possible to use just a standard AFM (atomic force microscopy) cantilever with a hole in the centre of the pyramidal tip.

As already said, the point source must be brought near the surface that will be imaged within nanometres and then must be scanned over the surface, without touching it, while the optical signal from the surface is collected and detected. There are two possible scanning procedures. In the first one, the probe is fixed at a mean distance from the surface. In the second operation the signal from the surface is kept constant.

The distance between the point light source and the sample surface is usually controlled through a feedback mechanism that is unrelated to the SNOM signal. Currently, most instruments use one of the following two types of feedback: *shear* force or *normal* force. In the shear force feedback, or tuning fork feedback, one uses the circumstance that the vibration of a probe in a direction parallel to the sample surface is influenced by the proximity of the sample. Typically, the probe is oscillating at the resonance frequency of its mechanical support so that

[20]Novotny L, Hafner (1994) Phys Rev. E50:4094

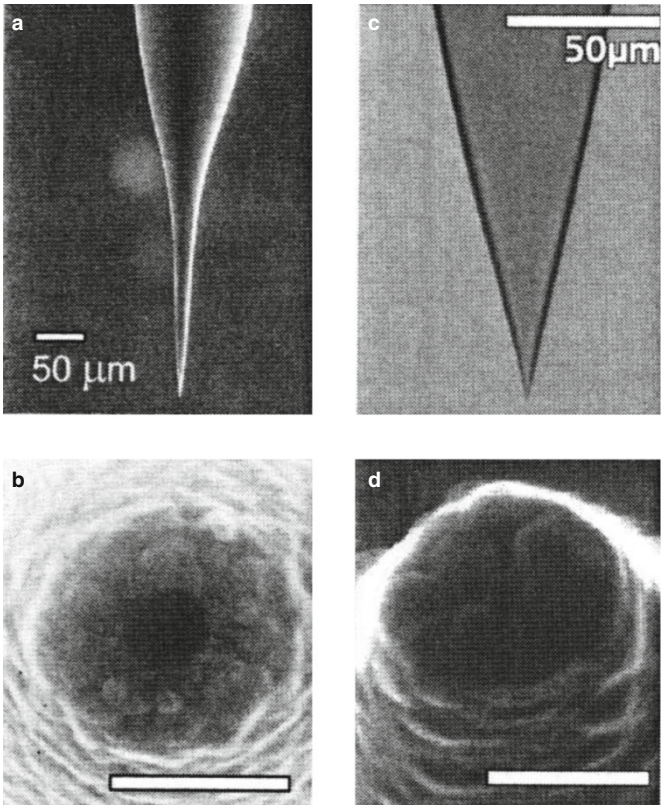


Fig. 4.4 Aluminium coated aperture probes prepared by pulling (**a**) and (**b**) and etching (**c**) and (**d**): (**a**) and (**c**) macroscopic shape, SEM and optical image. (**b**) and (**d**) SEM close up of the aperture region. Scale bar corresponds to 300 nm. (From B Hecht et al., J Chem Phys. 112, 7761 (2000)[21] reproduced with permission from AIP)

the straight tip is mounted to a tuning fork, which is then oscillated at its resonance frequency. The amplitude of this oscillation is strongly dependent on the tip-surface distance, and can be effectively used as a feedback signal.

This method has the advantage that any type of probe tip can be used as long as it is shaped approximately like a pencil and is small enough to be attached to a tuning fork. A different method is to use the normal force with a cantilever as in atomic force microscope.

Aperture-type near-field microscopy is limited in resolution because the effective diameter of an aperture cannot be smaller than twice the skin-depth of the metal used for coating the glass taper. The skin depth is between 6 and 10 nm for good metals at optical frequencies. In order to further push the confinement of light one

[21]Hecht B et al (2000) J Chem Phys. 112:7761

may use the fact that a near-field is created by any irradiated small material structure. One may use for this purpose small metallic object put near to the probe.

The main points in the working of a SNOM are

- the fabrication of near-field optical probes and the understanding of their working
- the knowledge of the field distribution emitted by the probe
- the tip-sample interaction and the image formation.

Near field microscopy remains primarily a surface inspection technique. However, no special sample preparation is needed and different kinds of samples can be used (conductive, non conductive, etc.).

Near field optical imaging is a sampling technique, i.e. the sample is probed point by point by raster-scanning the local probe over the sample surface and recording for every image pixel a corresponding optical signature. The local probe functions as an optical antenna, converting localized energy into radiation, and vice-versa.

Typically, a near-field optical microscope renders two simultaneously recorded images: a topographic image originating from keeping the shear-force feedback signal constant and an optical near-field image due to spatially varying optical properties of the sample and due to probe-sample distance variations. They correspond to the two operations already described: scanning the probe in a plane at a constant height from the mean surface plane or scanning maintaining a constant gap-width between the tip and the sample.

SNOM can be operated in many different ways. Most common today is the use of aperture probes for transmission microscopy, either in illumination or in collection. However, many samples or substances are opaque, so that working in reflection is necessary. The reflected light can be collected by optics close to the tip, or by the fibre probe itself.

In the transmission mode, the sample is illuminated through the probe, and the light passing through the sample is collected and detected. In the reflection mode, the sample is illuminated through the probe, and the light reflected from the sample surface is collected and detected. In the collection mode imaging, the sample is illuminated with a macroscopic light source from the top or bottom, and the probe is used to collect the light from the sample surface. It is also possible to use the probe for both the illumination of the sample and for the collection of the reflected light.

Using SNOM it is possible to see the near-field fluorescence image of a single molecule with a resolution in the range of 10 nm.[22]

4.3 Scanning Tunnelling Optical Microscope (STOM)

The *Scanning Tunnelling Optical Microscope*, also called *Photon Scanning Tunnelling Microscope* (PSTM) uses near-field detection and far-field illumination.

[22]Hosaka N, Saiki T (2001) J Microscop. 202:362

The use of an evanescent wave that is formed by total internal reflection (TIR) of a laser beam at the sample surface was introduced by Reddick[23] and independently by Courjon et al.[24] In a STOM the object is illuminated by an evanescent wave generated at the face of a prism or slide, and the field is detected via a fibre probe in the near zone of the sample. The intensity of the evanescent wave is probed by means of a sub-wavelength dielectric tip (a bare tapered glass fibre) that is dipped into the evanescent field and locally couples some of the light into the probe where it is converted into propagating modes that are guided towards a detector. A purely optical distance control mechanism is set up that is sensitive to local changes of the sample topography, the refractive index, and the polarization.[25] This concept is called Photon-Scanning Tunnelling Microscopy (PSTM or STOM) because of its similarity to the tunnelling of electrons through a small gap between a conducting surface and the probing tip. In this case are photons that tunnel from the sample to the high refractive index of the probe, through a thin layer of air. One of the major advantage of the STOM over the electron scanning tunnelling microscope is that an electrically conductive surface is no longer necessary. This makes imaging of biological samples much simpler and eliminate the need to coat the samples in gold or another conductive metal. Furthermore, the optical tunnelling microscope can be used to measure the optical properties of a sample allowing spectroscopic analysis and can be coupled with techniques such as photoluminescence, absorption and Raman spectroscopy.

Because the evanescent field is an exponential decaying function of the distance between the probe and the surface, the degree of the recorded intensity from the tip corresponds to the height of the probe from the sample surface and in many cases the signal is primarily modulated by the sample surface topography. However, the refractive index and absorption properties of the sample can cause further changes to the detected evanescent field, making it necessary to separate optical data from topographical data. This is often accomplished by coupling STOM to other techniques such as an atomic force microscopy.

The interpretation of results may therefore be not an easy task also in this type of microscope. The degree of optical coupling to the probe tip in practice is usually determined by calibration of the instrument, although detailed mathematical models have been published.[26]

A schematic of the set-up of a photon scanning tunnelling microscope is shown in Fig. 4.5, that describes an experiment in which standing evanescent waves are detected at the surface of a glass prism.

[23]Reddick RC et al (1989) Phys Rev. B39:767

[24]Courjon D et al (1989) Opt Comm. 71:23

[25]van Hulst NF et al (1992) Ultramicroscopy 42–44:416; Ferrell TL et al (1991) J Vac Sci Technol. B9:525

[26]Goumri-Said S et al (2005) Optics Comm. 244:245; Goumri-Said S et al (2004) Optical Quantum Electron 36:787

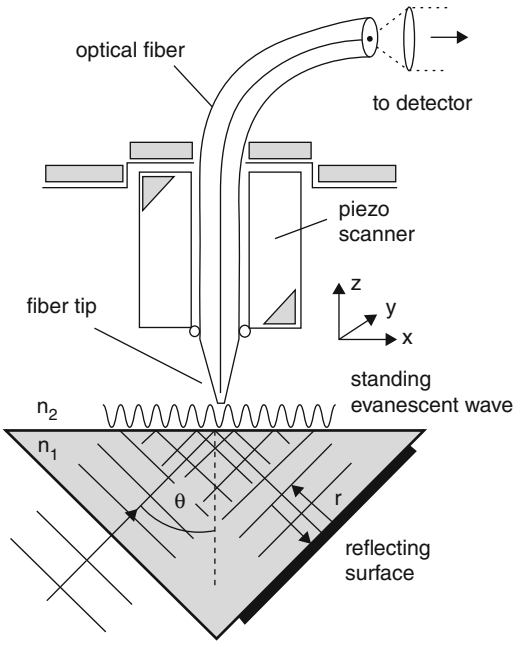


Fig. 4.5 The set-up of a PSTM for collecting light from a standing evanescent wave bound to the surface of a prism. (From A Meixner et al., Appl. Opt. 33, 7995 (1994) © Optical Society America)[27]

Two counter-propagating evanescent waves are formed by total reflection of a laser beam inside a 90° prism, which has a reflection coating on the output face. The probing tip is brought in a controlled fashion into the optical near field by means of a four-segment piezoelectric tube scanner. The tip scatters the evanescent field and creates propagating waves that can be detected in the far field by a fibre.

The amplitude of the evanescent wave at the boundary is given by the vector sum of the electric field amplitudes of the incident and reflected beams. A standing evanescent wave is formed by the interference of the two totally reflected counter-propagating laser beams of identical wavelengths and polarized perpendicular to the plane of incidence. The intensity profile of the standing evanescent wave is

$$I(x,y) = I_o \exp(-2z/d)\{1 + r^2 + 2r\cos[(2\pi/\Lambda)x + \varphi]\} \tag{4.3}$$

where

$$I_o = (c\varepsilon_o/2)|E_o|^2 = (c\varepsilon_o/2)\left[4n_1^2\cos^2\theta/(n_1^2 - n_2^2)\right]|E_i|^2 \tag{4.4}$$

being E_i and θ the amplitude of the incident wave, and incidence angle respectively; n_1 and n_2 the indices of refraction of the prism's glass and the air, respectively, and E_o the amplitude of the evanescent wave at the boundary. The intensity profile is modulated along the x direction with a period

$$\Lambda = \lambda o/(2n_1 \sin\theta) \tag{4.5}$$

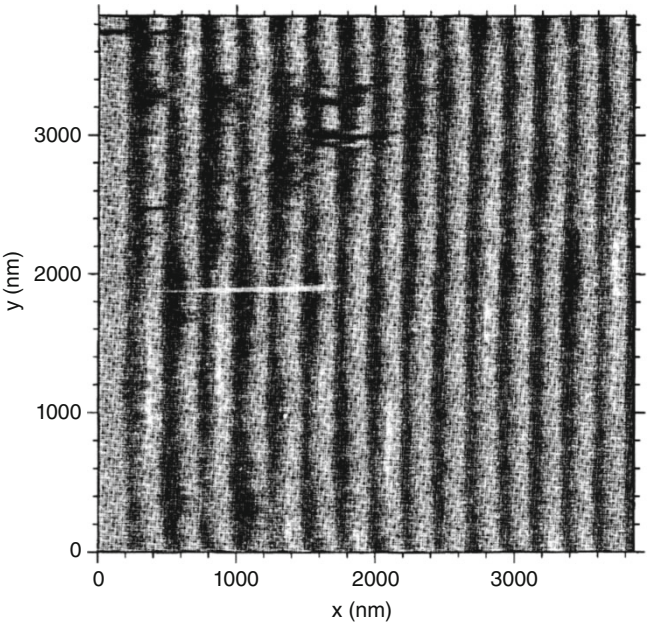


Fig. 4.6 Recorded pattern of the standing evanescent wave bound to the surface of the prism. (From A Meixner et al., Appl. Opt. 33, 7995 (1994) © Optical Society America)[27]

which is half of the wavelength of an individual evanescent wave and the modulation depth 2r is given by the ratio of the electric field components of the two counter-propagating waves; φ is an arbitrary phase shift.

Figure 4.6 shows the contour of the standing evanescent wave detected using a sharp transparent optical fibre that convert the evanescent field at its tip into a guided mode propagating along the fibre.[27]

In all scanning-probe modalities, the probe must be kept very near to the sample in order to obtain images in which structure may be recognized. Even then, the connection between the measured signal and the sample properties has proven to be problematic because of the interaction between the probe and the sample. Moreover when the probe is pulled away from the sample, the image becomes blurred because the field is not focused. However sub-wavelength tomography may be computed from data obtained at distances up to roughly a wavelength away from the sample depending on the level of noise.[28] The main point is that a STOM measurement is inherently holographic.[29] The sample is

[27]Meixner A, Bopp M, Tarrach G (1994) Appl Opt. 33:7995

[28]Scott Carney P et al (2004) Phys. Rev. Lett. 92:163903

[29]Bozhevolnyi SI, Vohnsen B (1996) Phys Rev Lett. 77:3351

illuminated by an evanescent field and the same field serves as a reference wave. The measurement is thus the near-field equivalent of Gabor hologram.[30] The hologram may be measured a large fraction of a wavelength from the sample, where the near field is normally out of focus. Carney et al.,[28] demonstrated that the data thus obtained, through not amenable to direct interpretation, provide enough information to determine quantitatively the two-dimensional structure of a thin sample.

STOM has found a number of applications in a number of fields as biology, optoelectronics and optical waveguides.

Balistreri et al.,[31] reported a technique for tracking a femtosecond laser pulse as it propagates through a photonic structure using STOM. The position in space of the propagating pulse at a particular point in time can be pinpointed and its group velocity can be measured.

The evanescent field of the pulse travelling inside the waveguide is picked up by a sub-wavelength probe. The signal is then interferometrically mixed with a laser pulse that is split of from the same laser source and has propagated along the reference branch of the setup. The length of the reference branch, and thus the time it takes the reference pulse to travel in this branch, is controlled with an optical delay line. The near field optical probe is scanned across the photonic structure while its height above the structure is kept constant. The pulses are launched into the structure, where they propagate. Heterodyne detection is established by acousto-optical modulation in the reference branch of the interferometer, enabling a separation of amplitude and phase information from the measured interference fringes with a dual-output lock-in amplifier. An optical delay line in the reference branch completes the pulse tracking setup. As a result, the centre of the pulse can be followed in time.

4.4 Total Internal Reflection Fluorescence (TIRF)

In cell and molecular biology, attention is especially given to the behaviour of cells very near to the surface of the sample. In conventional fluorescence microscopy, all the fluorescent molecules of the specimen are excited and detected and the contribution of the fluorophores bound to surface is often overwhelmed by the contribution of the fluorescence of all the molecules. In such microscopes the illumination is made by light incident at large angles so not to enter in the detected field. These large angles are however lower than the critical angle. The technique is called *epi (subcritical angle illumination) fluorescence microscopy*.

[30]Gabor D (1948) Nature 161:777

[31]Balistreri MLM et al (2001) Science 294:1080

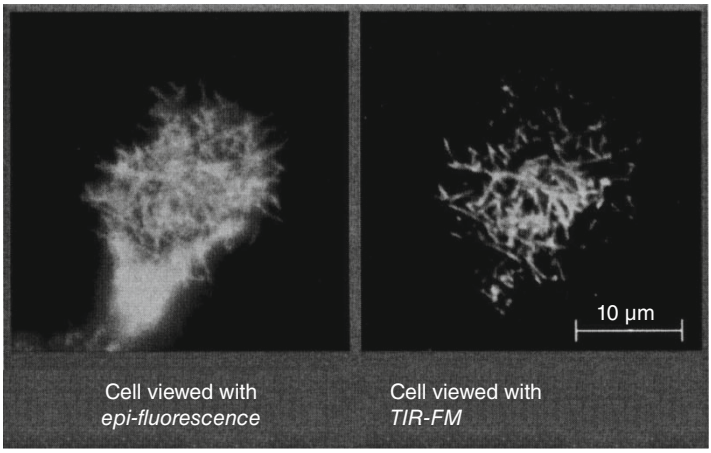


Fig. 4.7 Comparison between a cell viewed with a conventional fluorescence microscope (on the left) in which the illumination was made with an angle less than the critical angle (epi) and the same cell viewed with TIRF (right)

The idea of using total reflection so to illuminate the cells with the evanescent field was first described by E J Ambrose in 1956[32] and completed by D Axelrod[33] in 1981 who set the basis for the *Total Internal Reflection Fluorescence* (TIRF) microscopy, proposing to use the evanescent field to excite fluorescent molecules that sit immediately over the surface for a thickness of a fraction of wavelength. The idea of using evanescent waves had also been discussed in 1970 by H Nassenstein[34] in a different contest.

If fluorescent molecules dispersed in an aqueous solution sit over a glass support behind which a light beam is incident under the critical angle so to produce evanescent waves, only the molecules within around 100 nm of the interface are illuminated, and fluoresce, provided, of course, that they have radiative transitions at the wavelength bandwidth of the illuminating beam. As a consequence, the image observed with a microscope objective is sharper with respect to what happens with a conventional fluorescence microscope image that is blurred due to fluorescence from out of focus sources (Fig. 4.7).

The fluorescence emission near a dielectric surface is rather anisotropic. The polarization of the evanescent field depends on the incident light polarization, which can be either s or p.

For s-polarized incident light, the evanescent electric field vector remains purely normal to the plane of incidence. For p-polarization incident light, the

[32]Ambrose EJ (1956) Nature 178:1194

[33]Axelrod D (1981) J Cell Biology 89:141

[34]Nassenstein H (1970) Opt Comm. 2:231

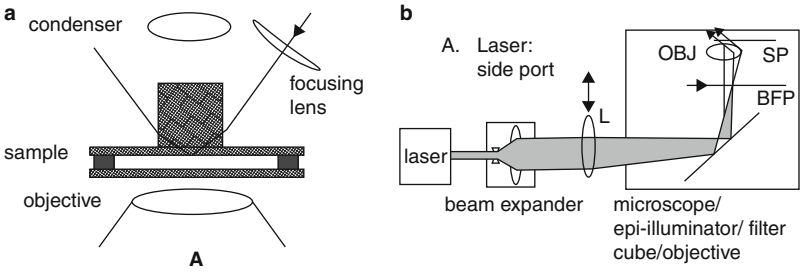


Fig. 4.8 (**a**) The prism disposition. (**b**) The objective lens method. (From D Axelrod, Methods Ezymology 361, 1 (2003) © Elsevier)

evanescent electric field vector direction remains in the plane of incidence, but it has a longitudinal component that is absent in the case of sub-critical refracted light.

The molecule emission depends on the orientation of its dipole with respect to the field components. The emission of a single molecule in an inhomogeneous environment is amply treated by Novotny and Hecht.[35]

There are essentially two configurations for TIRF: the *prism method*, and the *objective lens method*. In the prism technique (Fig. 4.8a), a focused laser beam produces the evanescent field at the base of a prism (or a cube) which is in contact with the sample that is viewed by the objective of a microscope. This technique has several limitations, primarily due to geometric constraints on specimen manipulation. Another disadvantage is that the evanescent field region is on the opposite side of the objective, requiring imaging of this region through the bulk of the specimen.

The objective lens technique uses an objective with large numerical aperture[36] in which the pump light is introduced at the periphery to ensure that light will exit at an angle greater than the critical angle. In Fig. 4.8b the beam from a laser is focused at the back focal plane of the objective OBJ at a radial position sufficient to lead to supercritical angle propagation into the sample SP. Moving the focusing lens L transversely changes the angle of incidence at the sample and allows switching between subcritical (EPI) and supercritical (TIR) illumination. Accomplishing evanescent field excitation through the use of high aperture objectives offers greater flexibility in specimen manipulation than with the prism method.[37]

[35]Novotny L, Hecht B (2006) Principles of nano-optics. Cambridge

[36]We remember that the numerical aperture of an objective gives the maximum angle of acceptance of light into it. The relation between the numerical aperture N.A. and the maximum acceptance angle θ is N.A. = nsinθ, where n is the refractive index and θ is one-half the objective angular aperture.

[37]Axelrod D (2003) Methods Enzymology 361:1

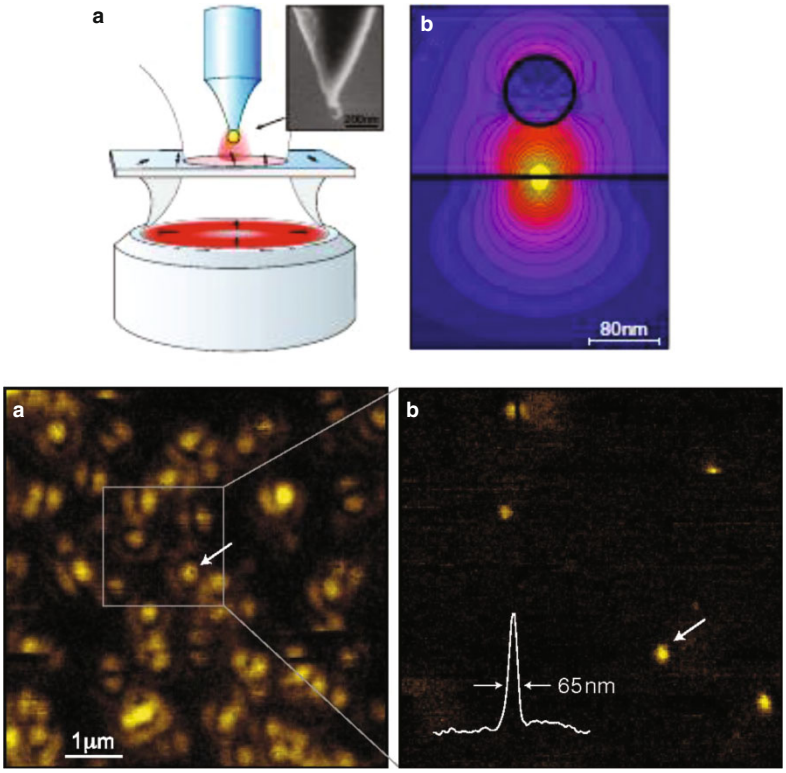


Fig. 4.9 Single molecule fluorescence imaging using the scattered light from a laser irradiated gold particle as a light source. Upper part (**a**) Experimental arrangement. An 80 nm gold particle is attached to the end of a etched glass tip and irradiated by a radially polarized laser beam. (**b**) Field distribution of an emitting dipole ($\lambda = 650$ nm) located 2 nm underneath the surface of a glass substrate and faced by a gold particle separated by a distance of $z = 60$ nm from the glass surface. Lower part Fluorescence images of a sample of nile blue molecules. (**a**) Confocal image using a radially polarized excitation beam. (**b**) Near-field image of the marked area exploiting the local field enhancement of a gold nanoparticle ($d = 80$ nm) positioned into the laser focus ($z = 5$ nm). (P Anger et al. Phys. Rev. Lett. 96, 113002 (2006) reproduced with permission from AIP)[29]

The study of fluorescent molecules can be made also using an aperture SNOM in which the molecule is illuminated by a small aperture and the fluorescent light is collected through the same aperture, filtering out the exciting wavelength. In this way imaging of single molecules was obtained with a resolution in the range of 10 nm.[38]

Alternatively single molecule fluorescence imaging can be obtained using the scattered light from a laser irradiating a gold particle as a light source[39] (Fig. 4.9).

[38]Hosaka N, Saiki T (2001) J Micros. 202:362

[39]Anger P et al (2006) Phys Rev Lett. 96:113002

The excitation of fluorescence as a function of the distance from the gold particle depends also on the orientation of the molecule dipole.

Using aperture-less microscopy to study single fluorescent molecules it is necessary to take into account the interaction between the sharp tip and the dipolar emitter which may affect the excited state lifetime and the fluorescence rate of single molecules.[40]

[40]Trabesinger W et al (2002) App Phys Lett. 81:2118; Trabesinger W et al (2003) J Micros. 209:249; Vickery JA, Dun RC (1999) Biophys J. 76:1812

Chapter 5
Plasmons

5.1 Introduction

An electron gas in a solid, as for example the free electrons of metals, has the capacity to undergo collective motions that have been named *plasma oscillations*.

These collective plasma oscillations are due to the long range nature of the Coulomb interaction between conductions electrons in metals, as was pointed out for the first time by Pines and Bohm[1] and discussed in detail by Pines.[2]

The presence of surfaces introduces new modes of plasma oscillations in addition to the bulk one. These new modes can be excited by incident electrons or photons and can be detected experimentally. They are also strongly dependent on the properties of the surfaces so that they can be used to make sensors. Moreover these surface oscillations allow to shrink optical signals in dimensions much smaller than their vacuum wavelength offering the way to have optical circuits of nanometric dimensions.

The quanta of these oscillations are called *plasmons*.[3] Plasmons are electromagnetic excitations coupled to electrons in a metal or doped semiconductor. There are several types of plasmons. We will start considering *bulk plasmons* that are longitudinal excitations in the volume of a metal not coupled with an external electromagnetic field, and *surface plasmon polaritons* that are electromagnetic excitations propagating at the interface between a dielectric and a conductor, evanescently confined in the perpendicular direction. These electromagnetic surface waves arise via the coupling of the electromagnetic field to oscillations of the conductor's electron plasma. The frequency ω of these longitudinal oscillations is tied to its wave vector k_x by a dispersion relation $\omega(k_x)$.

[1]Pines D, Bohm D (1952) Phys Rev. 85:338; Pines D (1956) Rev Mod Phys. 28:184

[2]Pines D (1963) Elementary excitations in solids, Benjamin, New York, chap. 3

[3]The name plasmon was given by Ritchie (1957), Phys Rev. 106:874. Later Stern EA, Ferrel RA (1960) Phys Rev. 120:130, coined the term surface plasmon

© Springer International Publishing AG 2017

M. Bertolotti et al., *Evanescent Waves in Optics*, Springer Series in Optical Sciences,
DOI 10.1007/978-3-319-61261-4_5

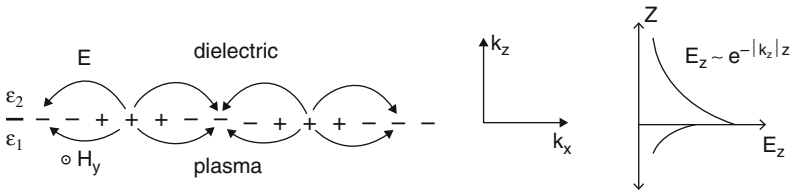


Fig. 5.1 A two dimensional representation of a plasmon polariton wave. On the left a visual representation of the electron oscillations, at centre the two involved k vector components, and on the right the behaviour of the electric field in the TM wave configuration. (From Raether 1988)

These charge fluctuations, which can be localized in the z direction within the Thomas-Fermi screening length of about 1 Å, are accompanied by a mixed transversal and longitudinal electromagnetic field which disappears at $z \to \infty$ (Fig. 5.1) and has its maximum in the surface $z = 0$, typical for surface waves. This explains their sensitivity to surface properties which allows to make very sensible sensors.

Similar coupling may also occur with the free electrons in a doped semiconductor. Also non-travelling plasmons can be generated as resonant oscillations near or inside metal nanoparticles.

Another kind of coupling may occur tying the electromagnetic wave to oscillations of the ions of the crystal lattice in ionic crystals or in polar materials. The quanta of these oscillations are called *plasmon polaritons*.

In addition, in nonlinear optical materials, the possibility to change the refractive index as a function of the optical power allows to generate new kinds of plasmons. Surface evanescent waves can be produced also at the interfaces between many different materials, like anisotropic, chiral, negative index materials, etc.

5.2 Plasmon Solutions

Combining the curl Maxwell's equations in the absence of external sources

$$\nabla \times \mathbf{E} = -\partial \mathbf{B}/\partial t; \quad \nabla \times \mathbf{H} = \partial \mathbf{D}/\partial t \tag{5.1}$$

leads to the wave equations in the case of nonmagnetic media

$$\nabla \times \nabla \times \mathbf{E} = -\mu_o \partial^2 \mathbf{D}/\partial t^2 \tag{5.2}$$

$$k(\mathbf{k} \cdot \mathbf{E}) - k^2 \mathbf{E} = -\varepsilon(k\omega)\omega^2/c^2 \mathbf{E} \tag{5.3}$$

in the time and Fourier domains, respectively. Here and in the following, ε indicates the relative dielectric constant.

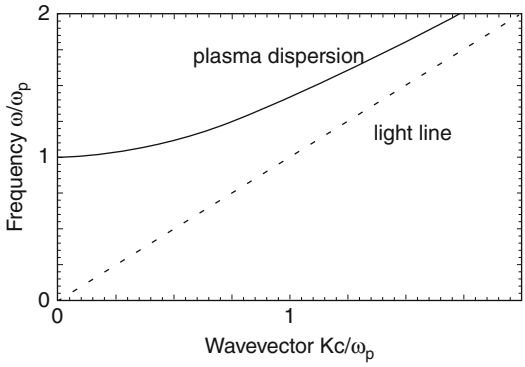


Fig. 5.2 The dispersion relation of the free electron gas. Electromagnetic wave propagation is only allowed for $\omega > \omega_p$. The dotted line represent the vacuum dispersion relation $\omega = kc$. (From S A Maier, Plasmonics Fundamentals and Applications, Springer 2007)

Depending on the polarization direction of the electric field vector two cases arises.

We may have transverse waves, if $\mathbf{k} \cdot \mathbf{E} = 0$, yielding the generic dispersion relation

$$k^2 = \varepsilon(k\omega)\omega^2/c^2. \tag{5.4}$$

In the transparent region of a metal, for $\omega > \omega_p$ the dispersion relation of travelling waves, neglecting damping, is

$$\varepsilon = 1 - \omega_p^2/\omega^2 \tag{5.5}$$

which put into (5.4) gives

$$k^2 = (\omega^2/c^2)(1 - \omega_p^2/\omega^2) \tag{5.6}$$

or

$$\omega^2 = \omega_p^2 + k^2c^2. \tag{5.7}$$

This relation is plotted for a generic free electron metal in Fig. 5.2.

The figure shows that for $\omega < \omega_p$ the propagation of transverse electromagnetic waves is forbidden inside the metal. For $\omega > \omega_p$ the plasma supports transverse propagation with group velocity

$$v_g = d\omega/dk < c. \tag{5.8}$$

The material is transparent. We have already discussed in Chap. 1 the modifications of (5.5) in metals due to interband absorptions.

5.3 Bulk Plasmons

Equation (5.3) allows also longitudinal waves, if $\mathbf{k} \cdot \mathbf{E} = k\mathbf{E}$. In this case the left hand side of (5.3) is zero and therefore a solution can exists only if $\varepsilon(k, \omega) = 0$, signifying that longitudinal collective oscillations can occur at frequencies corresponding to zeros of $\varepsilon(\omega)$.

When $\varepsilon(\omega) = 0$ (at the plasma frequency as shown by 1.38), $D = 0 = \varepsilon_o E + P$, and the electric field is a pure depolarization field $E = -P/\varepsilon_o$. Collective longitudinal oscillations of the conduction electron gas can be produced, which may be justified with the simple following argument.

In a slab, a collective displacement of the electrons by a distance x produces a surface charge density $\sigma = \pm nex$ (with n electron number density in the conductor) at the slab boundaries, which establishes a homogeneous field (Fig. 5.3)

$$E = nex/\varepsilon_o \tag{5.9}$$

inside the slab. This field act on the displaced electrons which experience a restoring force –neE, so that their movement can be discussed via the following equation

$$nmd^2x/dt^2 = -n^2e^2x/\varepsilon_o \tag{5.10}$$

or

$$d^2x/dt^2 + \omega_p^2 x = 0. \tag{5.11}$$

Equation (5.11) shows that the electron sea oscillates at its natural frequency that is the plasma frequency ω_p. The quanta of these charge oscillations are called *bulk (volume) plasmons*. Due to the longitudinal nature of the excitation, bulk plasmons do not couple to transverse electromagnetic waves and can only be excited by particle impact.

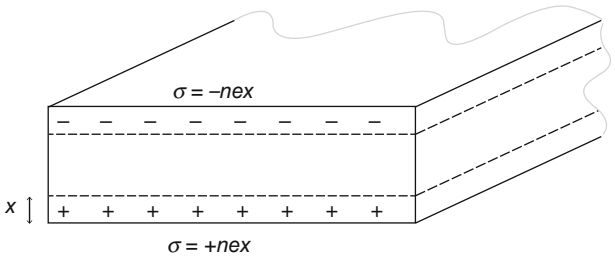


Fig. 5.3 Longitudinal collective oscillations of the conduction electrons of a metal: volume plasmons

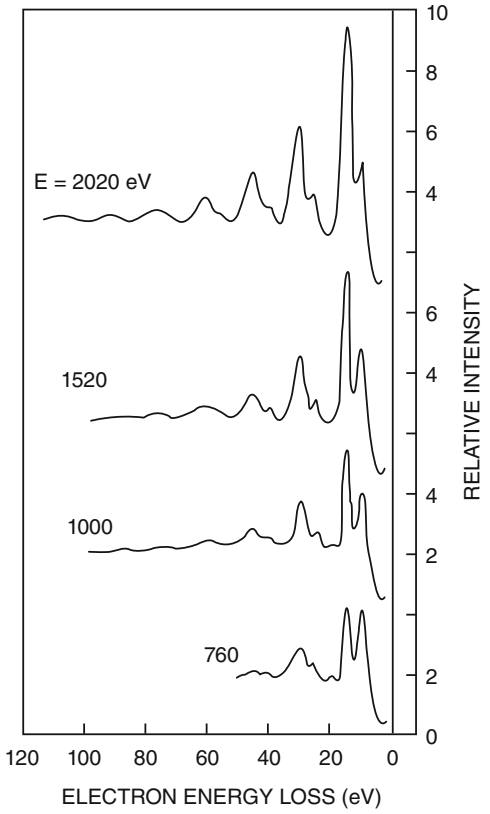


Fig. 5.4 Characteristics electron energy loss spectra of aluminium for primary electron energies of 760, 1000, 1520 and 2020 eV. No significant difference is found in the energy loss values for the different primary energies. The loss spectra are composed of combinations of elementary 10.3 and 15.3 eV energy losses which correspond to bulk and surface plasmon excitation in the Al. (From C J Powell and J B Swan, Phys. Rev. 115, 869 (1959) reproduced with permission from AIP)

It is possible to excite a bulk plasmon by electrons crossing a metallic film or reflected by it. The Coulomb field of the electron couples with the electrostatic fluctuations associated with the plasma oscillations. The electron reflected or transmitted shows an energy loss equal to integer multiples of the plasmon energy.[4] Figure 5.4 shows the energy distribution of electrons of various energies scattered by an Al specimen. The electron beam excited both bulk and surface plasmons.

5.4 Surface Plasmon Polaritons (SPPs)

Surface Plasmon Polaritons (SPPs) can be studied considering the curl Maxwell's equations. We look for solutions of the kind

$$E = E_o^{\pm} \exp[i(\beta x \pm k_z z - \omega t)] \tag{5.12}$$

[4]Powell CJ, Swan JB (1959) Phys Rev. 115:869; Powell CJ, Swan JB (1959) Phys Rev. 116:81

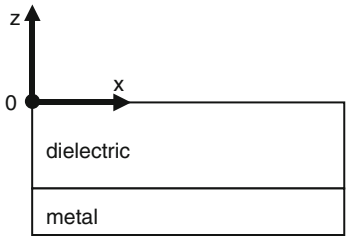


Fig. 5.5 Geometry for SPP propagation at a single interface between a metal and a dielectric

with + for $z > 0$, − for $z < 0$, β propagation constant of the wave, and with imaginary k_z, which causes the exponential decay of the field E_z for $z \to \pm\infty$.

The wavevector $\beta = k_x$ lies parallel to the x direction; $k_x = 2\pi/\lambda_p$, where λ_p is the wavelength of the plasma oscillation. The effective index is $n_{eff} = \lambda_o\beta/2\pi$, where λ_o is the vacuum wavelength.

Accordingly we write here in the case of harmonic time dependence ($\partial/\partial t = -i\omega$) considering a one-dimensional problem with the wave propagating in the x-direction (see Fig. 5.5)

$$\partial E_z/\partial y - \partial E_y/\partial z = i\omega\mu_o H_x \tag{5.13a}$$

$$\partial E_x/\partial z - \partial E_z/\partial x = i\omega\mu_o H_y \tag{5.13b}$$

$$\partial E_y/\partial x - \partial E_x/\partial y = i\omega\mu_o H_z \tag{5.13c}$$

$$\partial H_z/\partial y - \partial H_y/\partial z = -i\omega\varepsilon_o\varepsilon E_x \tag{5.13d}$$

$$\partial H_x/\partial z - \partial H_z/\partial x = -i\omega\varepsilon_o\varepsilon E_y \tag{5.13e}$$

$$\partial H_y/\partial x - \partial H_x/\partial y = -i\omega\varepsilon_o\varepsilon E_z. \tag{5.13f}$$

For plane waves $E(x, y, z) = E(z)e^{i\beta x}$, where $\beta = k_x$ is the propagation constant of a wave propagating along the x-direction ($\partial/\partial x = i\beta$) and homogeneity in the y-direction ($\partial/\partial y = 0$), the equations simplify as

$$\partial E_y/\partial z = -i\omega\mu_o H_x \tag{5.14a}$$

$$\partial E_x/\partial z - i\beta E_z = i\omega\mu_o H_y \tag{5.14b}$$

$$i\beta E_y = i\omega\mu_o H_z \tag{5.14c}$$

$$\partial H_y/\partial z = i\omega\varepsilon_o\varepsilon E_x \tag{5.14d}$$

$$\partial H_x/\partial z - i\beta H_z = -i\omega\varepsilon_o\varepsilon E_y \tag{5.14e}$$

$$i\beta H_y = -i\omega\varepsilon_o\varepsilon E_z. \tag{5.14f}$$

The system of (5.14) can be split into two set of independent equations describing TM or p and TE or s modes, respectively.

For TM modes we have

$$\partial E_x/\partial z - i\beta E_z = i\omega\mu_o H_y \tag{5.15a}$$

$$\partial H_y/\partial z = i\omega\varepsilon_o\varepsilon E_x \tag{5.15b}$$

$$i\beta H_y = -i\omega\varepsilon_o\varepsilon E_z. \tag{5.15c}$$

For TE modes we have

$$\partial E_y/\partial z = -i\omega\mu_o H_x \tag{5.16a}$$

$$i\beta E_y = i\omega\mu_o H_z \tag{5.16b}$$

$$\partial H_x/\partial z - i\beta H_z = -i\omega\varepsilon_o\varepsilon E_y. \tag{5.16c}$$

Considering TM waves, we derive the wave equation for the magnetic field

$$\partial^2 H_y/\partial z^2 + (k_o^2\varepsilon - \beta^2)H_y = 0 \tag{5.17}$$

and the two components of the electric field are

$$E_x = -(i/\omega\varepsilon_o\varepsilon)\partial H_y\partial z \tag{5.18a}$$

$$E_z = -(\beta/\omega\varepsilon_o\varepsilon)H_y. \tag{5.18b}$$

The general solution for H_y is

$$H_y = A_1\exp(k_i z) + A_2\exp(-k_i z) \tag{5.19}$$

with

$$k_i^2 = -(k_o^2\varepsilon_i - \beta^2) \tag{5.20}$$

and i = d for the dielectric or vacuum and m for the metal. The solutions for $z > 0$ taking into account the condition that the solution should go to zero for $z \to \infty$ (therefore $A_1 = 0$) are

$$H_y(z) = A_2 e^{i\beta x}\exp(-k_d z) \tag{5.21a}$$

$$E_x(z) = iA_2(1/\omega\varepsilon_o\varepsilon_d)k_d e^{i\beta x}\exp(-k_d z) \tag{5.21b}$$

$$E_z(z) = -A_2(\beta/\omega\varepsilon_o\varepsilon_d)e^{i\beta x}\exp(-k_d z). \tag{5.21c}$$

where

$$k_d^2 = \beta^2 - k_o^2 \varepsilon_d \tag{5.22}$$

The solutions for $z < 0$ are

$$H_y(z) = A_1 e^{i\beta x} \exp(ik_m z) \tag{5.23a}$$

$$E_x(z) = -iA_1 (1/\omega\varepsilon_o\varepsilon_m) k_1 e^{i\beta x} \exp(ik_m z) \tag{5.23b}$$

$$E_z(z) = -A_1 (\beta/\omega\varepsilon_o\varepsilon_m) e^{i\beta x} \exp(ik_m z) \tag{5.23c}$$

where

$$k_m^2 = \beta^2 - k_o^2 \varepsilon_m. \tag{5.24}$$

Continuity of H_y and E_x at $z = 0$ gives

$$A_2 = A_1 \tag{5.25}$$

$$k_d/k_m = -\varepsilon_d/\varepsilon_m. \tag{5.26}$$

Using (5.26), (5.22) and (5.24), one obtains

$$\beta = k_o [\varepsilon_m \varepsilon_d/(\varepsilon_m + \varepsilon_d)]^{1/2}. \tag{5.27}$$

This expression is valid for both real and complex dielectric constants, i.e. for conductors without and with attenuation. We are looking for interface waves that propagate along the surface. This requires real β. Looking at (5.27) this can be fulfilled if the sum and the product of the dielectric functions are either both positive or both negative. The wavevector components along z (k_m and k_d) must be imaginary and both positive. The two requirements are fulfilled if $\varepsilon_m\varepsilon_d < 0$ and $\varepsilon_m + \varepsilon_d < 0$ with the negative part larger than the positive one. Therefore the surface wave can exist if $\varepsilon_m < 0$ with $|\varepsilon_m| > |\varepsilon_d|$.

If we assume, besides a real ω and ε_d, a complex ε_m with $\varepsilon_m'' < |\varepsilon_m'|$ we obtain a complex $\beta = \beta' + i\beta''$ with

$$\beta' = (\omega/c) \left[\varepsilon_m' \varepsilon_d/(\varepsilon_m' + \varepsilon_d)\right]^{1/2} \tag{5.28a}$$

and

$$\beta'' = (\omega/c) \left[\varepsilon_m' \varepsilon_d/(\varepsilon_m' + \varepsilon_d)\right]^{1/2} \left[\varepsilon_m''/2(\varepsilon_m')^2\right] \tag{5.28b}$$

the imaginary part β'' determines the internal absorption.

The penetration depths at which the electric field falls to 1/e in the dielectric and the metal are defined $\delta_d = 1/k_d$ and $\delta_m = 1/k_m$, respectively. From (3.22) and (5.24) they are

$$\delta_d = (1/k)|(\varepsilon_d + \varepsilon_m)/(-\varepsilon_d^2)|^{1/2} \tag{5.29a}$$

$$\delta_m = (1/k)|(\varepsilon_d + \varepsilon_m)/(-\varepsilon_m^2)|^{1/2}. \tag{5.29b}$$

The dielectric constant of a metal is usually greater than that of a dielectric, so the SPP field has a shorter penetration depth in a metal than in the dielectric.

An idea of the numerical values can be obtained considering, for ex., silver on silicon oxide at a wavelength of 620 nm. It is $\varepsilon_d = 2.127$, $\varepsilon_m = -15.037 + i1.017$, $\delta_d = 167$ nm and $\delta_m = 23.6$ nm. At 400 nm wavelength ($\varepsilon_d = 2.160$, $\varepsilon_m = -3.773 + i0.675$) it is $\delta_d = 39.0$ nm and $\delta_m = 22.0$ nm.

Let us now see what can be said for TE waves. We have

$$\partial E_y/\partial z = -i\omega\mu_o H_x \tag{5.30a}$$

$$i\beta E_y = i\omega\mu_o H_z \tag{5.30b}$$

$$\partial H_x/\partial z - i\beta H_z = -i\omega\varepsilon_o\varepsilon E_y. \tag{5.30c}$$

This time we obtain a wave equation for the electric field

$$\partial^2 E_y/\partial z^2 + (k_o^2\varepsilon - \beta^2)E_y = 0 \tag{5.31}$$

and the relations for the magnetic field

$$H_x = (i/\omega\mu_o)\partial E_y\partial z \tag{5.32a}$$

$$H_z = (\beta/\omega\mu_o)E_y. \tag{5.32b}$$

The solution for E_y which should go to zero per $z \to \pm\infty$ is

$$E_y = A_1\exp(k_i z) + A_2\exp(-k_i z) \tag{5.33}$$

with

$$k_i^2 = -(k_o^2\varepsilon_i - \beta^2) \tag{5.34}$$

and i = d for dielectric or vacuum and m for the metal. The solutions for $z > 0$ are

$$E_y(z) = A_2 e^{i\beta x}\exp(ik_d z) \tag{5.35a}$$

$$H_x(z) = -iA_2(1/\omega\mu_o)k_d e^{i\beta x}\exp(ik_d z) \tag{5.35b}$$

$$H_z(z) = A_2(\beta/\omega\mu_o)e^{i\beta x}\exp(ik_d z) \tag{5.35c}$$

where

$$k_d^2 = \beta^2 - k_o^2\varepsilon_d. \tag{5.36}$$

The solutions for $z < 0$ are

$$E_y(z) = A_1 e^{i\beta x}\exp(ik_m z) \tag{5.37a}$$

$$H_x(z) = iA_1(1/\omega\mu_o)k_m e^{i\beta x}\exp(ik_m z) \tag{5.37b}$$

$$H_z(z) = A_1(\beta/\omega\mu_o)e^{i\beta x}\exp(ik_m z) \tag{5.37c}$$

where

$$k_m^2 = \beta^2 - k_o^2\varepsilon_m. \tag{5.38}$$

Continuity of E_y and H_x at $z = 0$ gives

$$A_2 = A_1 \tag{5.39a}$$

$$-iA_2 k_d/\omega\mu_o = iA_1 k_m/\omega\mu_o. \tag{5.39b}$$

The continuity conditions then require

$$A_1 = A_2 \tag{5.40a}$$

and

$$k_m + k_d = 0. \tag{5.40b}$$

Since the confinement to the surface requires $\mathrm{Re}[k_m] > 0$ and $\mathrm{Re}[k_d] > 0$ the second relation cannot be fulfilled. No surface plasmon polariton exist for TE polarization.

5.5 Properties of Plasmons

Figure 5.6 shows a plot of (5.27) for a metal with negligible damping described by the real Drude dielectric function (5.5) for an air ($\varepsilon_d = 1$) interface. Due to their bound nature, the SPP excitations correspond to the part of the dispersion curve lying to the right of the light line. Between the regime of bound and radiative modes, a frequency gap region with purely imaginary β prohibiting propagation exists.

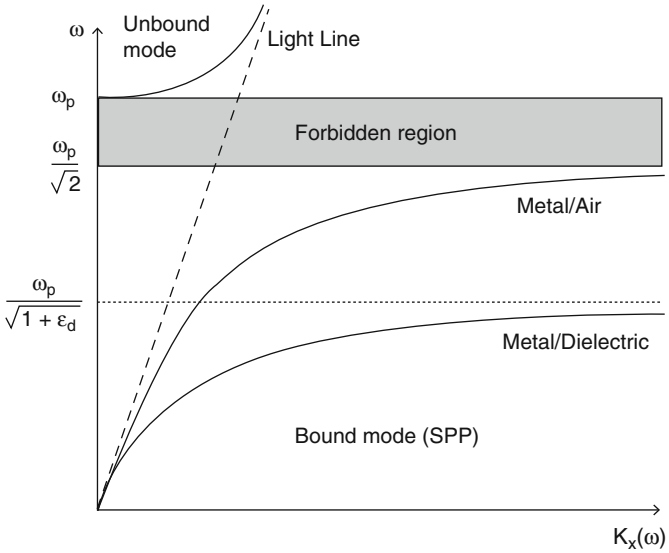


Fig. 5.6 Plasmon polariton dispersion relation in the absence of damping. The slanting line represents the light cone inside which a wave is propagating and outside which is evanescent (surface wave). The free propagation is described in the upper left side (unbound mode). Two PPS corresponding to metal/air and metal/dielectric are shown as a function of beta (K_x)

For small wave vectors corresponding to low (mid-infrared or lower) frequencies, the SPP propagation constant is close to k_o the free space wave vector, at the light line, and the waves extend over many wavelengths into the dielectric space. In this regime, SPPs therefore acquire the nature of a grazing-incident light field and are also known as Sommerfeld-Zenneck waves.[5]

In the opposite regime of large wave vectors, the frequency of the SPPs approaches the characteristic *surface plasmon frequency*

$$\omega_{sp} = \omega_p / (1 + \varepsilon_d)^{1/2} \tag{5.41}$$

as can be shown by inserting the free-electron dielectric function (5.5).

In the limit of negligible damping of the conduction electrons oscillation (implying $\mathrm{Im}[\varepsilon_m(\omega)] = 0$), the wave vector $k_x = \beta \to \infty$ as the frequency approaches ω_{sp}, and the group velocity $v_g \to 0$. The mode acquires electrostatic character, and is known as the *surface plasmon*.

Real metals however, suffer both from free-electron and interband damping. Therefore, $\varepsilon_m(\omega)$ is complex, and so is the SPP propagation constant β. The

[5]Zenneck J (1907) Ann Phys. 23:846; Goubau G (1950) J Appl Phys. 21:1119. Zenneck surface waves are essentially surface plasmons at radio frequencies, see Hill DA, Wait JR (1978) Radio Science 13:969

travelling SPPs are damped with an energy attenuation length (also called propagation length) $L = (2\mathrm{Im}[\beta])^{-1}$, typically between 10 and 100 μm in the visible regime, depending upon the metal/dielectric configuration.

Compared with the dispersion relation of completely un-damped SPPs, the propagation constant of the SPPs no more goes to infinity but approaches now a maximum, finite wave vector at the surface plasmon frequency ω_{sp}.

SPPs at a silver/air interface, at 450 nm for example, have a propagation length $L = 16$ μm and an energy confinement in the air in the z direction $\delta_d = (\beta^2 - \varepsilon_d^2)^{-½} = 180$ nm. At 1.5 μm however, $L = 1080$ μm and $\delta_d = 2.6$ μm. The better the confinement, the lower the propagation length. Field confinement below the diffraction limit of half the wavelength in the dielectric can be achieved close to the plasma frequency.

The amplitude of normal and tangential electric field components in the dielectric and metal are linked by

$$E_z^d = i[-(\varepsilon_m/\varepsilon_d)]^{½} E_x^0 \quad \text{and} \quad E_z^m = -i[-(\varepsilon_d/\varepsilon_m)]^{½} E_x^0. \tag{5.42}$$

Metals typically have $|\varepsilon_m| > \varepsilon_d$, and therefore the normal component is dominant in the dielectric and the tangential component is dominant in the metal. SPP damping occurs due to ohmic losses in the metal, and these are due to the longitudinal electric field component.

An important feature of surface plasmons is that for a given energy $\hbar\omega$ the wave vector k_x is always larger than the wave vector of light in free space (Fig. 5.6). The light line is asymptotically approached by the SPP dispersion for small energies. The physical reason for the increased momentum of the SPP is the strong coupling between light and surface charges. The light field has to "drag" the electrons along the metal surface. Consequently, this means that a SPP on a plane interface cannot be excited by light of any frequency that propagates in free space. Excitation of a SPP by light is only possible if a wave vector component of the exciting light can be increased over its free-space value.

The localization and concomitant field enhancement offered by SPPs at the interface between a conductor and a dielectric with refractive index n is due to a large SPP propagation constant $\beta > k_o n$, leading to evanescent decay of the field perpendicular to the interface. The amount of confinement increases with β according to (5.22) and (5.24). Conversely, localization significantly decreases for frequencies $\omega < \omega_p$, where β goes to $k_o n$.

Finally we want to note that in the limit of a perfect conductor, the internal fields are identically zero and thus perfect metals do not support electromagnetic surface modes, forbidding the existence of SPPs.

However Pendry and co-workers have shown that bound electromagnetic surface waves mimicking SPPs can be sustained even by a perfect conductor, provided that its surface is periodically corrugated.[6]

[6]Pendry JB et al (2004) Science 305:847

One way to study SPPs is to use a SNOM with uncoated optical tip. In this way it is possible to probe the surface polariton field directly over the surface.[7] The fibre tip is usually scanned at a fixed distance above the local sample surface within a few tens of nanometers, obtaining simultaneous topographical and optical imaging which allow to correlate the optical field distribution over the surface with its topography.

In this way propagation and scattering of SPPs can be studied.[8]

5.6 Excitation and Coupling of Plasmons

Plasmons cannot be coupled or excited directly by an electromagnetic wave because the momentum conservation cannot be fulfilled directly.

The wave vector β of the plasmon, according to the dispersion (5.27), exceeds that of the free radiation at the same frequency in the inactive medium; that is $\beta > k$, where k is the wave vector of light on the dielectric side of the interface. In other words the plasmon phase velocity at the interface is less than the respective phase velocity in the inactive medium (for ex. vacuum). Therefore, the projection along the interface of the momentum $k_x = k\sin\theta$ of photons impinging under an angle θ to the surface normal is always smaller than the SPP propagation constant β, even at grazing incidence, prohibiting phase-matching. The situation is similar to the one we encountered with the problem of coupling external radiation in a slab waveguide.

As in that case it is necessary to devise a way to fulfil the matching condition. One may couple the external field with the plasmon using evanescent waves by means of a grating, fibre and waveguide couplers or some roughness on the surface. In the case of grating the parallel component of the wave vector of the wave, falling onto the surface, increases by the wave vector connected with the Fourier components of the surface roughness with the periodicity of the grating.

This grating coupling is shown schematically in Fig. 5.7. The p radiation falling onto the grating excites the plasmon wave if the angle of incidence onto the grating satisfies the condition

$$k_x = \omega(\varepsilon_2)^{1/2}\sin\theta + 2\pi m/a \tag{5.43}$$

where a is the period of the grating and m is an integer.

The coupling may also take place through evanescent wave coupling by using a prism. A p polarized beam reflected at the interface between a dielectric of higher dielectric constant ε, usually in the form of a prism (Fig. 5.8), and the metal will have an in-plane momentum $k_x = k(\varepsilon)^{1/2}\sin\theta$, which is sufficient to excite

[7]Pohl DW, Courjon D (eds) (1993) Near field optics, Amsterdam

[8]Zayats AV, Smolyaninv II (2003) J Opt A Pure Appl Opt. 5:S16

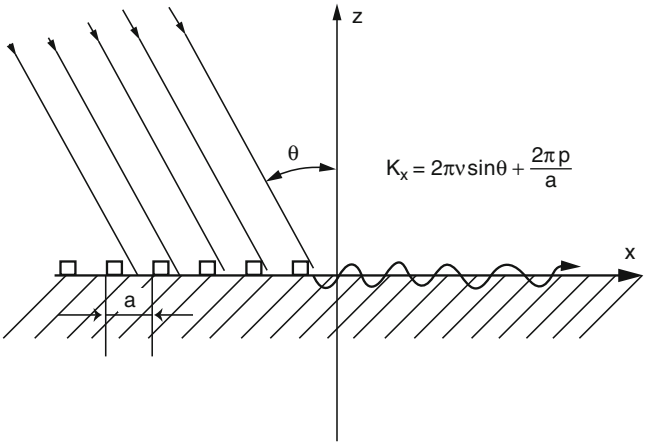


Fig. 5.7 Grating coupling of external radiation to SSP. (From Agranovich, fig. 4, p. 107)

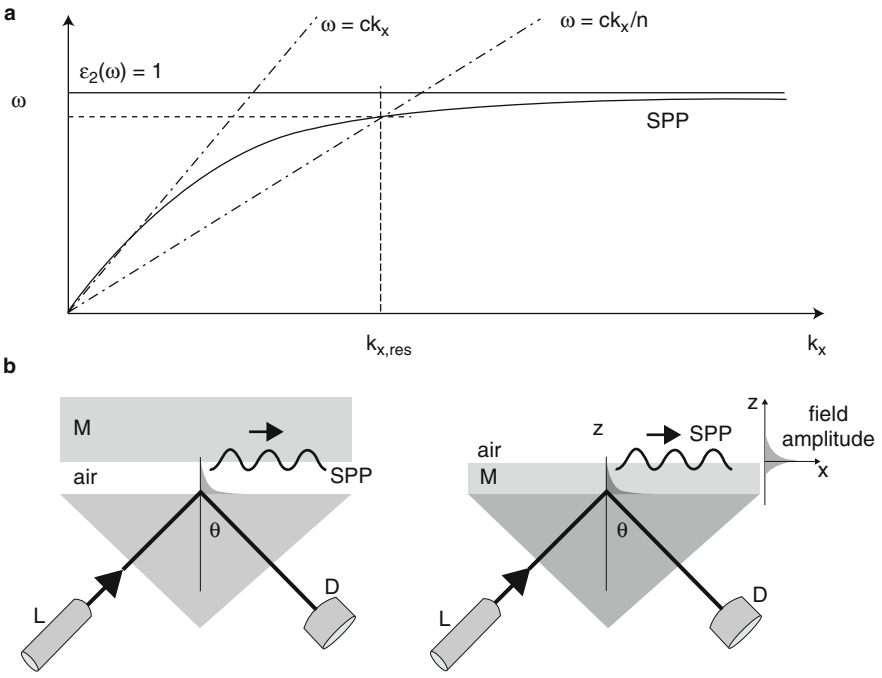


Fig. 5.8 **a** Prism coupling and dispersion; **b** the two most used geometries: the Otto (left) and Kretschmann (right) configuration. (From L Novotny and B Hecht, Principles of Nano-Optics, Cambridge 2006) reproduced with permission by Cambridge University Press

SPPs at the interface between the metal and the lower index dielectric, i.e. in this case at the metal/air interface (Fig. 5.8).

This coupling scheme, also known as attenuated total internal reflection (ATR), involves tunnelling of the fields of the excitation beam to the metal/air interface where SPP excitation takes place.

Two different geometries Kretschmann (Fig. 5.8 right) and Otto (Fig. 5.8 left) are employed. The most common configuration is the Kretschmann method[9] in which a thin metal film is evaporated on top of a glass prism. Photons from a beam impinging from the glass side at an angle greater than the critical angle of total internal reflection tunnel through the metal film and excite SPPs at the metal/air interface.

A different geometry is the Otto configuration[10] (Fig. 5.8 left) in which the prism is separated from the metal film by a thin air gap. Total internal reflection takes place at the prism/air interface, exciting SPPs via tunnelling to the air/metal interface. This configuration is preferable when direct contact with the metal surface is undesirable or when the metal thickness is too large to allow effective tunnelling of photons.

In the Otto configuration the tail of an evanescent wave at a glass/air interface is brought into contact with a metal/air interface that supports SPPs. For a sufficiently large separation between the two interfaces (gap width) the evanescent wave is only weakly influenced by the presence of the metal. By tuning the angle of incidence of the totally reflected beam inside the prism, the resonance condition for the excitation of SPPs, i.e. the matching of the parallel wave vector components, can be fulfilled. The excitation of a SPP will show up as a minimum in the reflected light in an experiment as described in Fig. 5.9.

The reflectivity of the system as a function of the angle of incidence and the gap width is shown in Fig. 5.10. A clear resonance is observed in gold at an angle of incidence of 43.5°. For a too small gap width the resonance is broadened and shifted due to radiation damping of the SPP. This is caused by the presence of the glass half-space, which allows the SPP to couple back in the glass and rapidly decay radiatively by transforming the evanescent SPP field into a propagating field in the glass. For a gap width that is too large the SPP can no longer be efficiently excited and the resonance vanishes.

Similar considerations apply for the Kretschmann configuration (Fig. 5.11). Also in this case if the metal is too thin, the SPP will be strongly damped because of radiation damping into the glass. If the metal is too thick, the SPP can no longer be efficiently excited due to absorption in the metal. The figure shows the reflectivity of the excitation beam as a function of the metal thickness and the angle of incidence. As before, the resonant excitation of surface plasmons is characterized by a dip in the reflection curves.

The maximum possible field enhancement on a smooth surface is

$$|E_{SP}^{(z=0)}/E_{light}|^2 = (2/\varepsilon_i)(|Re(\varepsilon_m)|^2/Im(\varepsilon_m)[a/(1+|Re(\varepsilon_m)|)]) \tag{5.44}$$

[9]Kretschmann E, Raether H (1968) Z Naturf. 23A:2135

[10]Otto A (1968) Z Phys. 216:398

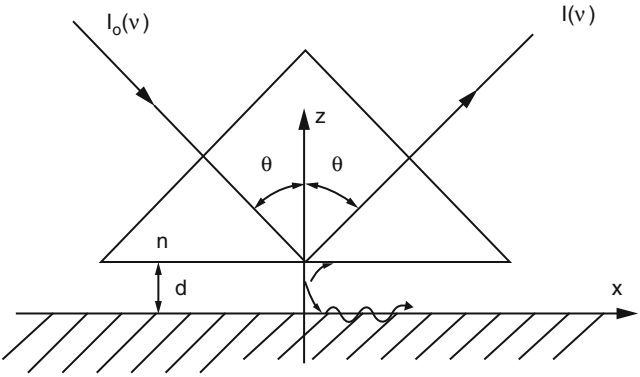


Fig. 5.9 In the prism coupling geometry a portion of the incident light is coupled to the SPP and a portion is reflected out of the prism as shown in the right part of the figure

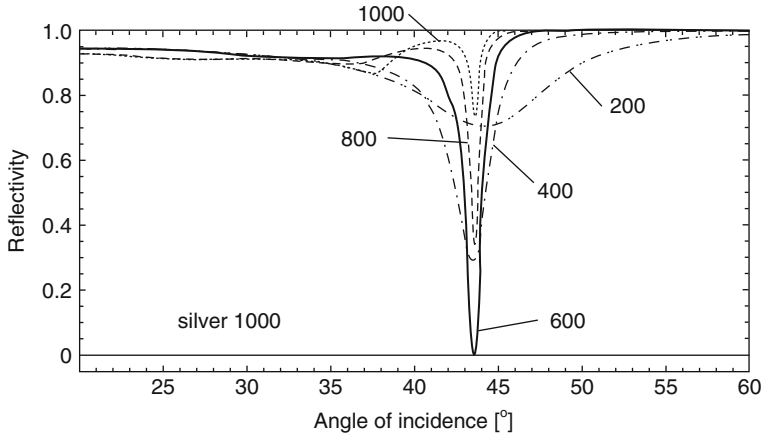


Fig. 5.10 Excitation of surface plasmons in the Otto configuration. The reflectivity of the exciting beam is plotted as a function of the incident angle and for different air gaps (in nm). The curves are evaluated for a gold film. (From L Novotny and B Hecht, Principles of Nano-Optics, Cambridge 2006) reproduced with permission by Cambridge University Press

where

$$a^2 = |\mathrm{Re}(\varepsilon_m)|(\varepsilon_s - 1) - \varepsilon_s,$$

E_{light} is the electric field of the incident light, ε_i, ε_m and ε_s are the dielectric constant of the adjacent medium, metal, and of the prism through which the metal film is illuminated to excite SPP, respectively. For a 60 nm thick silver film illuminated with red light, the intensity enhancement can reach more than two orders.

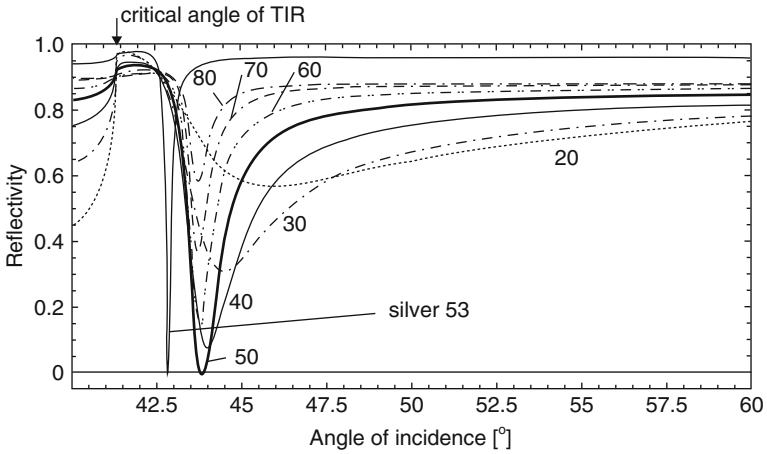


Fig. 5.11 Excitation of surface plasmons in the Kretschmann configuration. The reflectivity of the exciting beam is plotted as a function of the incident angle and for different air gaps (in nm). The curves are evaluated for gold film. For comparison a single trace is also plotted for silver. Due to the smaller damping of silver as compared with gold, it is much sharper. (From L Novotny and B Hecht, Principles of Nano-Optics, Cambridge 2006) reproduced with permission by Cambridge University Press

The parallel components of the wavevector necessary for plasmon excitation can also be found if the metal surface is rough[11] or are present suitably designed metallic particles. The scattered field wave-vector can match with the field component of SPP waves so that coupling to SPP can be accomplished locally.[12] On a rough surface, the SPP excitation conditions can be achieved without any special arrangement. Diffraction of light on surface features can provide coupling to the SPP modes because the diffracted components of light with all wave-vectors are present. However, the conditions for excitation are hill defined and the light coupling is not efficient.

Figure 5.12 shows excitation with a SNOM probe or by diffraction on surface features.

SNOM excitation provides the possibility to excite SPP locally at a given place on the surface at the SNOM tip position.

Whatever configuration is used, the excitation light should have a component of the electric field perpendicular to the metal surface or parallel to the propagation direction of the excited SPPs.

Prism coupling can be used either to couple external radiation to the SSP or to couple the SSP to the external radiation (Fig. 5.13).

[11]Raether H (1988) Surface plasmons on smooth and rough surfaces and on gratings. Springer-Verlag, Berlin

[12]Raether H (1982), Surface plasmons and roughness. In Agranovich V, Mills DL (eds) Surface polaritons, North-Holland

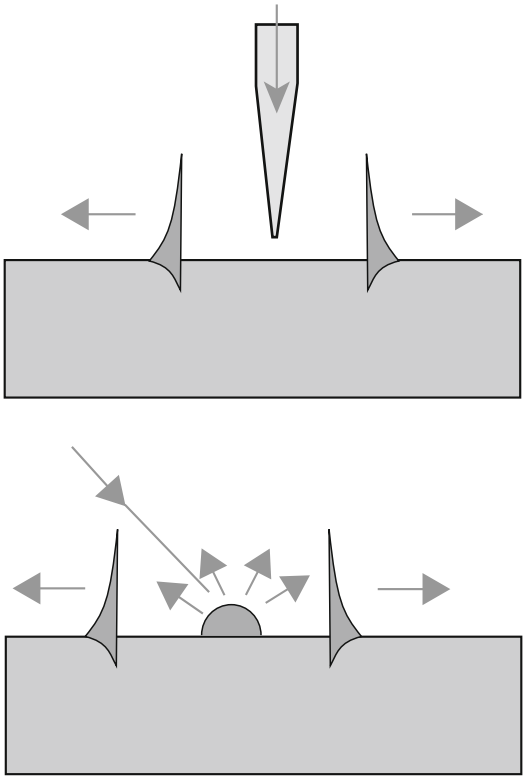


Fig. 5.12 Excitation of plasmons with a SNOM probe (upper) or by diffraction on a small particle (lower)

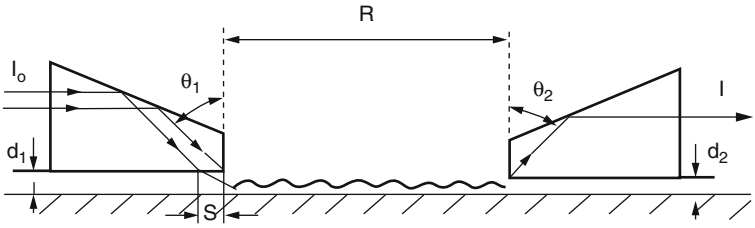


Fig. 5.13 Configuration of the coupling-decoupling prisms

5.7 Multilayer Systems

We consider a multilayer consisting of alternating conducting and dielectric films. In such a system, each single interface can sustain bound SPPs. When the separation between adjacent interfaces is comparable or smaller than the decay length z' of the interface mode, interaction between SPPs gives rise to coupled modes.

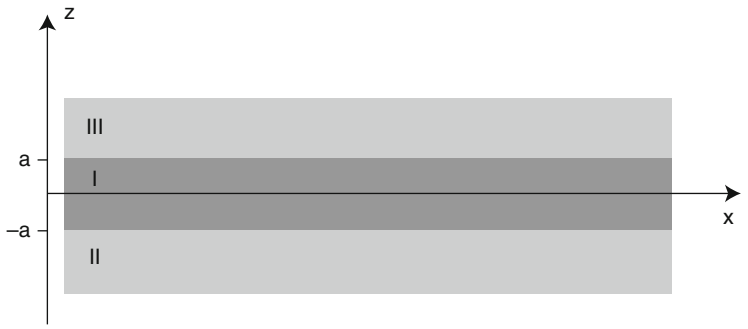


Fig. 5.14 Geometry of a three-layer system consisting of a thin layer I sandwitched between two infinite half spaces II and III

We consider a simple three-layer system (Fig. 5.14) in which a thin metallic layer (I) is sandwitched between two (infinitely) thick dielectric claddings (II, III).

Considering TM modes, we have, for $z > a$

$$H_y(z) = Ae^{i\beta x}\exp(-k_3 z) \tag{5.45a}$$

$$E_x(z) = iA(1/\omega\varepsilon_o\varepsilon_3)k_3 e^{i\beta x}\exp(-k_3 z) \tag{5.45b}$$

$$E_z(z) = -A(\beta/\omega\varepsilon_o\varepsilon_3)e^{i\beta x}\exp(-k_3 z) \tag{5.45c}$$

while for $z < -a$

$$H_y(z) = Be^{i\beta x}\exp(k_2 z) \tag{5.46a}$$

$$E_x(z) = -iB(1/\omega\varepsilon_o\varepsilon_2)k_2 e^{i\beta x}\exp(k_2 z) \tag{5.46b}$$

$$E_z(z) = -B(\beta/\omega\varepsilon_o\varepsilon_2)e^{i\beta x}\exp(k_2 z). \tag{5.46c}$$

We demand that the field decay exponentially in the cladding (II) and (III).

In the core region $-a < z < a$, the modes localize at the bottom and top interface couple, yielding

$$H_y(z) = Ce^{i\beta x}\exp(k_1 z) + De^{i\beta x}\exp(-k_1 z) \tag{5.47a}$$

$$E_x(z) = -iC(1/\omega\varepsilon_o\varepsilon_1)k_1 e^{i\beta x}\exp(k_1 z) + iD(1/\omega\varepsilon_o\varepsilon_1)k_1 e^{i\beta x}\exp(-k_1 z) \tag{5.47b}$$

$$E_z(z) = C(\beta/\omega\varepsilon_o\varepsilon_1)e^{i\beta x}\exp(k_1 z) + D(\beta/\omega\varepsilon_o\varepsilon_1)e^{i\beta x}\exp(-k_1 z). \tag{5.47c}$$

The continuity of H_y and E_x gives

$$\begin{aligned} A\exp(-k_3 a) &= C\exp(k_1 a) + D\exp(-k_1 a) \\ A(1/\varepsilon_3)k_3\exp(-k_3 a) &= -C(1/\varepsilon_1)k_1\exp(k_1 a) + D(1/\varepsilon_1)k_1\exp(-k_1 a) \end{aligned} \tag{5.48}$$

at $z = a$, and

$$\begin{aligned} B\exp(-k_2 a) &= C\exp(-k_1 a) + D\exp(k_1 a) \\ -B(1/\varepsilon_2)k_2\exp(-k_2 a) &= -C(1/\varepsilon_1)k_1\exp(-k_1 a) + D(1/\varepsilon_1)k_1\exp(k_1 a) \end{aligned} . \tag{5.49}$$

At $z = -a$. H_y further has to fulfil the wave equation in the three regions with

$$k_i^2 = \beta^2 - k_o^2\varepsilon_i \tag{5.50}$$

for $i = 1, 2, 3$. Solving the system of four equations results in

$$e^{-4k1a} = [(k_1/\varepsilon_1) + (k_2/\varepsilon_2 k_1/\varepsilon_1) + (k_3/\varepsilon_3)]/[(k_1/\varepsilon_1) - (k_2/\varepsilon_2 k_1/\varepsilon_1) - (k_3/\varepsilon_3)]. \tag{5.51}$$

Assuming the layer I and III to be equal in terms of their electrical properties, i.e. $\varepsilon_2 = \varepsilon_3$ and $k_2 = k_3$, the dispersion relation can be split into a pair of equations

$$\tanh k_1 a = -k_2\varepsilon_1/k_1\varepsilon_2 \tag{5.52a}$$

$$\tanh k_1 a = -k_1\varepsilon_2/k_2\varepsilon_1. \tag{5.52b}$$

The first equation describes modes of odd vector parity (E_x is odd, H_y and E_z are even functions with respect to the film mid-plane), while the other describes modes of even parity (E_x is even function, H_y and E_z are odd).

Figure 5.15 shows the dispersion relations of the coupled odd and even modes for an air/aluminium film/air. The odd modes have frequencies ω_+ higher than the respective frequencies for a single interface SPP, and the even modes a lower frequencies ω_-.

For large wave vectors β the limiting frequencies are

$$\omega_+ = \left[\omega_p/(1+\varepsilon_2)^{1/2}\right]\left[1 + 2\varepsilon_2\exp(-2\beta a)/(1+\varepsilon_2)\right]^{1/2} \tag{5.53a}$$

$$\omega_- = \left[\omega_p/(1+\varepsilon_2)^{1/2}\right]\left[1 - 2\varepsilon_2\exp(-2\beta a)/(1+\varepsilon_2)\right]^{1/2}. \tag{5.53b}$$

Odd modes have the interesting property that upon decreasing the metal film thickness, the confinement of the coupled SPP to the metal film decreases as the mode evolves into a TEM mode supported by the homogeneous dielectric environment. For real, absorptive metal described via a complex $\varepsilon(\omega)$, this implies a

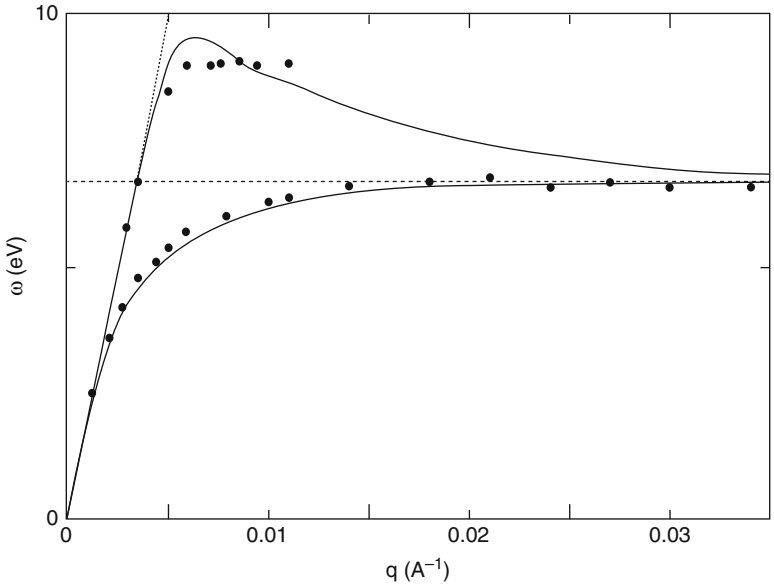


Fig. 5.15 Dispersion relation of the coupled odd and even modes for an air/Al/air structure with a metal core thickness 12 nm (dashed gray curves) surrounded by dielectric layers of equal thickness 4 nm. Solid line is theory with $\varepsilon_2 = 1$, $\varepsilon = 4$. The metal has $\omega_p = 15$ eV. The circles are experimental results by Petit et al.[13] The dashed line is the Plasmon frequency $\omega_p/(5)^{1/2}$ and the dotted line is $\omega = ck$. (From Pitarke et al.[14])

drastically increased SPP propagation length.[15] These plasmons are called *long-ranging plasmon polaritons*[16] *(LRSPP)*.

The even modes exhibit the opposite behaviour: their confinement to the metal increases with decreasing metal thickness, resulting in a reduction of propagation length (*short ranging Plasmon polaritons SRSPP*).

For example, in the Ag/SiO_2 system, for a 20 nm thick metal film and 1550 nm wavelength, the mode power attenuation coefficients of the long range SPP and the short range SPP are 2.8 cm^{-1} and 103 cm^{-1}, respectively.

Excitation of the LRSPP can be done with the prism configuration using p polarized light changing the angle of incidence. If the reflectance is plotted versus the incidence angle, then dips appear for each excited mode. Figure 5.16 shows the dips corresponding to the excitation of a couple of symmetric and anti-symmetric modes.

[13]Petit RB, Silcox J, Vincent R (1975) Phys Rev. B11:3116

[14]Pitarke JM et al (2007) Rep Progr Phys. 70:1

[15]Sarid D (1981) Phys Rev Lett. 47:1927

[16]See for example, Berini P (2009) Adv Optics Phot. 1:484, for a general introduction

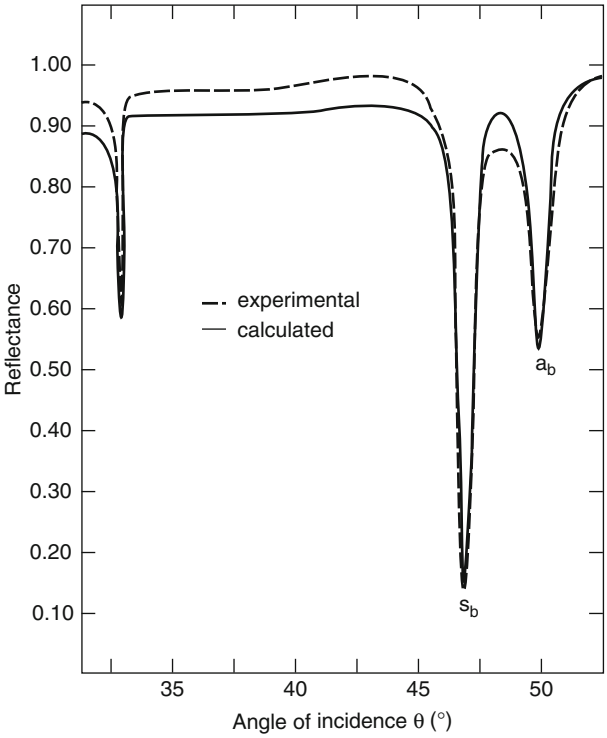


Fig. 5.16 Reflectance of p-polarized light at 632.8 nm vs angle θ of incidence at the base of the prism for an Ag film of 49 nm bounded by cryolite layers. Dips at 47° and 50° are due to coupling with the two modes, respectively. The dip at 33° is due to coupling with a TM guided mode in the bottom cryolite cladding. (From P Berini[17], Adv. Optics 1, 484 (2009) © Optical Society America)

A characterization of LRSPP has been made by Berini et al.[18] Due to the long range of propagation – of the order of a centimetre – these long-range surface plasmon polaritons have found many applications in nonlinear interactions, molecular scattering, fluorescence, surface enhanced Raman spectroscopy, etc.

Now we may consider the other possibility of a metal/dielectric/metal structure with $\varepsilon_2 = \varepsilon_2(\omega)$ the dielectric function of the metal and ε_1 the dielectric constant of the insulating core. Considering a symmetric configuration, when two identical SPP modes start overlapping with each other for small layer thickness, the propagation constants of the symmetric and anti-symmetric (individual) mode combinations become different and within a certain range of parameters, even a third

[17]Berini P (2009) Adv Optics Phot. 1:484

[18]Berini P (2005) J Appl Phys. 98:043109

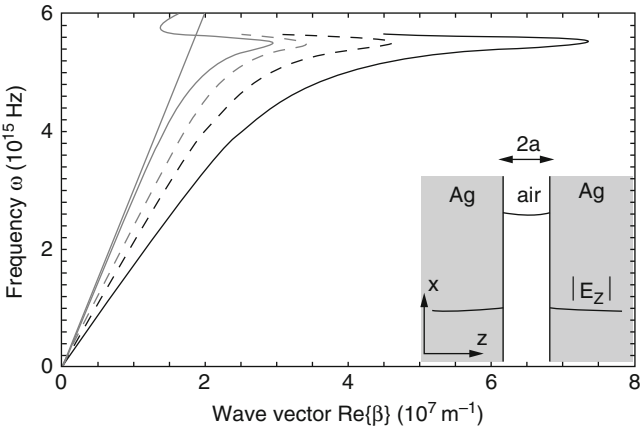


Fig. 5.17 Dispersion relation of the fundamental coupled SPP modes of a silver/air/silver multilayer geometry for an air core of size 100 nm (broken gray curve), 50 nm (broken black curve) and 25 nm (continuous black curve). Also shown is the dispersion of a SPP at a single silver/air interface (gray curve) and the air light line (gray line). (From S A Maier, Plasmonics Fundamentals and Applications, Springer 2007)

backward mode can propagate. The most interesting mode is the fundamental odd mode of the system, which does not exhibits a cut-off for vanishing core layer thickness.[19] This mode is called gap SPP (G-SPP) and has an odd symmetry of the longitudinal electric-field component E_x and an even symmetry of the transverse field component E_z. Figure 5.17 shows the dispersion relation of this odd mode for a silver/air/silver hetero-structure. This time, the dielectric function $\varepsilon(\omega)$ was taken as complex from ref.[20]

Thus β does not go to infinity as the surface plasmon frequency is approached, but fold back and eventually crosses the light line, as for SPPs propagating at single interfaces.

Large propagation constants β can be achieved even for excitation well below ω_p, provided that the width of the dielectric core is chosen sufficiently small. The ability to access such large wave vectors and thus small penetration lengths into the metallic layers by adjusting the geometry, indicates that localization effects, that for a single interface can only be sustained at excitations near ω_p, can also be obtained for excitation out in the far infrared. The G-SPP exploit in the most efficient way the dielectric gap between the metal walls, minimizing the absorption losses.[21]

[19]Prade B et al (1991) Phys Rev. B44:13556

[20]Johnson PB, Christy RW (1972) Phys Rev. B6:4370

[21]Bozhevolnyi SI et al (2007) Appl Phys. A89:225; Bozhevolnyi SI (2006) Opt Express 14:9467

5.8 Localized Surface Plasmons

Until now we have considered SPPs which are propagating, dispersive electromagnetic waves coupled to the electron plasma of a conductor at a dielectric interface. Localized surface plasmons (LSP) on the other hand are non-propagating excitations of the conduction electrons of metallic nanostructures coupled to the electromagnetic field. Small particles have electromagnetic resonances similar to the surface plasmon.[22]

These modes arise naturally from the scattering problem of a small, sub-wavelength conductive nanoparticle in an oscillating electromagnetic field. The curved surface of the particle exerts an effective restoring force on the driven electrons, so that a resonance can arise, leading to field amplification both inside and in the near-field zone outside the particle. This resonance is called *localized surface plasmon*. Plasmon resonances can be excited by direct light illumination.

When the particle is small compared to the wavelength of an incident plane wave, a plasmon can be excited that has the symmetry of a time varying dipole. Once excited the dipolar plasmon can, of course, radiate. For a sphere this resonance occurs at the frequency ω_R, for which the condition $\mathrm{Re}[\varepsilon(\omega_R)] = -2\varepsilon_o$ is satisfied.[23]

A sphere has other resonances whose frequencies may be obtained from the relation $\mathrm{Re}[\varepsilon(\omega_N)] = -[(N + 1)/N]\varepsilon_o$, where N is an integer. Except the $N = 1$ resonance, which is dipolar, the others have symmetries corresponding to higher multipoles and therefore cannot radiate.

For small spheres, only the dipole active excitation ($N = 1$) is important. With an increase in the size of the sphere, the contributions of higher multipoles become more and more significant, resulting in the long-wavelength limit (a very large sphere) for the surface plasmon frequency of a flat surface ($N \to \infty$). For ellipsoidal particles, in general three localised surface plasmon frequencies correspond to three major axes of an ellipsoid.[24] The range of frequencies where localized plasmon polariton modes exist is limited by the bulk plasma frequency.

LSP can be resonantly excited with light of appropriate frequency (and polarization) irrespective of the excitation light wave vector. Therefore, LSPs also effectively decay with light emission. LSPs can decay in surface polaritons and in turn can be excited by SPP. Since LSPs are confined to the particle (or a curved surface), this results in a significant electromagnetic field enhancement at small metallic particles due to the small volume of the LPS mode.[25] This effect contributes to numerous phenomena such as surface enhanced Raman scattering and second harmonic generation, and finds applications in aperture-less scanning near field microscopes.

---

[22]See for example, Maier SA (2007) Plasmonics. Springer; Zhang J, Zhang L (2012) Adv Opt Photon. 4:157; Zayats AV, Smolyaninov II (2003) J Opt A Pure Appl Opt. 5:S16

[23]For a discussion of the field in metal nano-particles, see for example, Campbell SD, Ziolkowski RW (2012) Adv Optoelectron. article ID368786 doi: 10.1155/2012/368786

[24]Kreibig U, Vollmer M (1995) Opical properties of metal clusters. Springer, Berlin

[25]Moskovits M (1985) Rev Mod Phys. 57:783

In the case of voids in the bulk of a metal, LSP frequencies can be found by replacing ε by 1/ε in the respective Dirichlet problem. The LSP frequencies of a particle and a void of the same shape are related to each other as

$$\omega_{\text{particle}}^2 + \omega_{\text{void}}^2 = \omega_p^2. \tag{5.54}$$

The spectrum of LSP associated with an ensemble of metallic particles (or voids) is determined by the interaction between the individual LSP resonances.

Also cylinders which have two significantly different characteristic sizes in different directions (f.e. the diameter $d \ll \lambda$ so that the electrostatic approximation can be used, and the length $L \gg \lambda$) have a discrete spectrum of resonances.

An emitting molecular dipole placed near a metal sphere can excite resonances of any order, provided the sphere is large compared to the wavelength of the radiation in question and assuming a fair degree of resonance between the dipole's frequency and those of the electromagnetic resonance of the sphere. Since only the dipolar plasmon can radiate, the fraction of the dipole's energy transferred to the metal sphere that has excited resonances corresponding to $N > 1$ will be degraded to heat. Likewise, an emitting dipole placed near a flat surface can excite a surface plasmon, since there is no momentum conservation restriction limiting the coupling of a point emitter and the plasmon. The plasmon, once excited, cannot radiate, however.

Let us consider the example of a small spherical metal particle of radius a illuminated with a homogeneous constant field of amplitude E_o directed along the x direction (Fig. 5.18) in the quasi-static near-field approximation that is valid provided $a \ll \lambda$.[26] We start considering the Laplace equation written in spherical coordinates (r, θ, φ)

$$(1/r^2 \sin\theta)\left[\sin\theta\partial(r^2\partial/\partial r)/\partial r + \partial(\sin\theta\partial/\partial\theta)/\partial\theta + (1/\sin\theta)\partial^2/\partial^2\varphi\right]\Phi(r,\theta,\varphi) = 0 \tag{5.55}$$

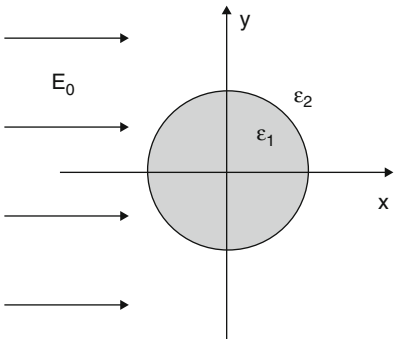


Fig. 5.18 Geometry of the excitation of a spherical particle by a plane wave

[26]See for example, Sernelius Bo E (2001), Surface modes in physics, Wiley-VCH, Berlin

with solutions of the form

$$\Phi(r,\theta,\varphi) = \sum \left[A_l r^l + B_l r^{-(l+1)}\right] P_l(\cos\theta) \tag{5.56}$$

where $P_l(\cos\theta)$ are the Legendre polynomials of order l. Due to the requirement that the potential remains finite at the origin, the solutions for the potentials Φ_{in} inside and Φ_{out} outside the sphere can be written as

$$\Phi_{in}(r,\theta) = \sum A_l r^l P_l(\cos\theta) \tag{5.57a}$$

$$\Phi_{out}(r,\theta) = \sum \left[B_l r^l + C_l r^{-(l+1)}\right] P_l(\cos\theta). \tag{5.57b}$$

The coefficients A_l, B_l and C_l are determined by the boundary conditions at $r \to \infty$ and at the sphere surface $r = a$. The requirement that $\Phi_{out} \to -E_o x = -E_o r\cos\theta$ as $r \to \infty$ requires that $B_l = -E_o$ and $B_l = 0$ for $l \neq 1$. The remaining coefficients A_l and C_l are defined by the boundary conditions at $r = a$. Continuity of the tangential electric field and of the normal component of the electric displacements at the surface of the sphere imply that

$$[\partial\Phi_{in}/\partial\theta]_{r=a} = [\partial\Phi_{out}/\partial\theta]_{r=a} \tag{5.58a}$$

$$\varepsilon_1[\partial\Phi_{in}/\partial r]_{r=a} = \varepsilon_2[\partial\Phi_{out}/\partial r]_{r=a}. \tag{5.58b}$$

We obtain $A_l = C_l = 0$ for $l \neq 1$, and finally we have

$$\Phi_{in} = -E_o[3\varepsilon_2/(\varepsilon_1 + 2\varepsilon_2)]r\cos\theta \tag{5.59a}$$

$$\Phi_{out} = -E_o r\cos\theta + E_o[(\varepsilon_1 - \varepsilon_2)/(\varepsilon_1 + \varepsilon_2)]a^3\cos\theta/r^2. \tag{5.59b}$$

Note that the potential is independent from φ. The field can be calculated from $E = -\nabla\boldsymbol{\Phi}$ and turns out to be

$$\mathbf{E_{in}} = E_o[3\varepsilon_2/(\varepsilon_1 + 2\varepsilon_2)](\cos\theta\mathbf{u_r} - \sin\theta\mathbf{u}_\theta) = E_o[3\varepsilon_2/(\varepsilon_1 + 2\varepsilon_2)]\mathbf{u_x} \tag{5.60a}$$

$$\mathbf{E_{out}} = E_o(\cos\theta\mathbf{u_r} - \sin\theta\mathbf{u}_\theta) + E_o[(\varepsilon_1 - \varepsilon_2)/(\varepsilon_1 + \varepsilon_2)]a^3/r^3(2\cos\theta\mathbf{u_r} + \sin\theta\mathbf{u_\theta}). \tag{5.60b}$$

Note that the field inside the sphere is homogeneous instead to be zero as it should be in a metal particle. Consequently the quasi-static approximation is valid only if the skin depth is larger that the radius a. The second term in (5.60b) is identical to the electrostatic field of a dipole μ located at the centre of the sphere, induced by the external field E_o. $\mu = \varepsilon_2\alpha(\omega)E_o$, with α denoting the polarizability

$$\alpha(\omega) = 4\pi\varepsilon_o a^3[(\varepsilon_1(\omega) - \varepsilon_2)/(\varepsilon_1(\omega) + \varepsilon_2)]. \tag{5.61}$$

The polarizability experiences a resonant enhancement under the condition that $|\varepsilon_1(\omega)+\varepsilon_2|$ is a minimum, which for the case of small or slowly-varying Im[ε] around the resonance simplifies to

$$\mathrm{Re}[\varepsilon(\omega)] = -2\varepsilon_2. \tag{5.62}$$

This relation is called the Froehlich condition and the associated mode (in an oscillating field) the *dipole surface plasmon* of the metal nanoparticle. For a sphere consisting of a Drude metal with dielectric function (1.34) in air, the Froehlich criterion is met at the frequency $\omega_o = \omega_p/(3)^{½}$.

The scattering cross section of the sphere is obtained by dividing the total radiated power of the sphere's dipole by the intensity of the exciting plane wave

$$\sigma_{scat} = (k^4/6\pi\varepsilon_o)|\alpha(\omega)|^2 \tag{5.63}$$

with k being the wave-vector in the surrounding medium. The resonance condition of a particle plasmon depends therefore sensitively on the dielectric constant of the environment. Thus, a gold or silver particle can be used as a sensing element since its resonance will shift upon local dielectric changes. The resonance for the silver particles is in the ultraviolet spectral range while for gold the maximum scattering occurs around 530 nm. A red shift of the resonance is observed if the dielectric constant of the environment is increased.

The absorption cross section is

$$\sigma_{abs} = (k/\varepsilon_o)\mathrm{Im}[\alpha(\omega)]. \tag{5.64}$$

It turns out that σ_{abs} scales with a^3 whereas σ_{scatt} scales with a^6. Consequently for large particles, extinction is dominated by scattering whereas for small particles it is associated with absorption. This effect can be used to detect extremely small metal particles down to 2.5 nm diameter which are used as labels in biological samples. The transition between the two size regimes is characterized by a distinct colour change. For example, small gold particles absorb green and blue light an thus render a red colour. On the other hand, larger gold particles scatter predominantly in the green and hence render a greenish colour.

The peculiar scattering properties of nanoparticles are well illustrated by the famous ancient Roman Lycurgus cup (Fig. 5.19) that when illuminated with white light from the behind shows a red colour, while when illuminated from the front appears green. The effect is due to the interplay between scattering and absorption.

The resulting electric dipole on the particle represents a restoring force and hence the nanoparticle can be considered in a first approximation as a harmonic oscillator, driven by a light wave and damped by some losses such as ohmic loses (essentially the production of heat) and radiative (scattering) losses that are equivalent to the re-emission of a photon on the expenses of nanoparticle plasmon excitation. Only light with a wavelength in resonance with an eigenmode of the nanoparticle plasmon oscillation is able to excite the plasmon. Therefore the nanoparticle plasmon

Fig. 5.19 Ancient Roman Lycurgus cup exhibited at the British Museum. Illuminated by a white light from behind looks red while seen from the front is green. The effect is due to embedded gold/silver alloy particles in the glass

manifest itself as a pronounced extinction band and as a strongly wavelength depending scattering.

Gustave Mie solved Maxwell's equations for the case of an incoming plane wave interacting with a spherical particle. The electromagnetic field is expanded in multipole contributions and the expansion coefficients are found by applying the correct boundary conditions for the electromagnetic fields at the interface between the metallic nanoparticle and its surrounding.[27] For very small spherical particles it is sufficient to consider only the first term of the expansion, which is the dipolar term.

In the Rayleigh limit (very small particles) the scattering, extinction, and absorption cross-sections of a spherical nanoparticle are given as

$$\sigma_{sca} = (3/2\pi)(\omega/c)^4 \varepsilon_{diel}^2 V^2 \left[(\varepsilon'_{met} - \varepsilon_{diel})^2 + (\varepsilon''_{met})^2\right] / \left[(\varepsilon'_{met} + 2\varepsilon_{diel})^2 + (\varepsilon'')^2\right] \tag{5.65a}$$

$$\sigma_{ext} = 9(\omega/c)\varepsilon_{diel}^{3/2}\varepsilon''_{met} / \left[(\varepsilon'_{met} + 2\varepsilon_{diel})^2 + (\varepsilon'')^2\right] \tag{5.65b}$$

$$\sigma_{abs} = \sigma_{ext} - \sigma_{sca} \tag{5.65c}$$

[27] Kreibig U, Vollmer M (1995) Optical properties of metal clusters. Springer; Bohren C, Huffmann D (1983) Absoption and scateringof light by small particles. John Wiley & Sons, New York

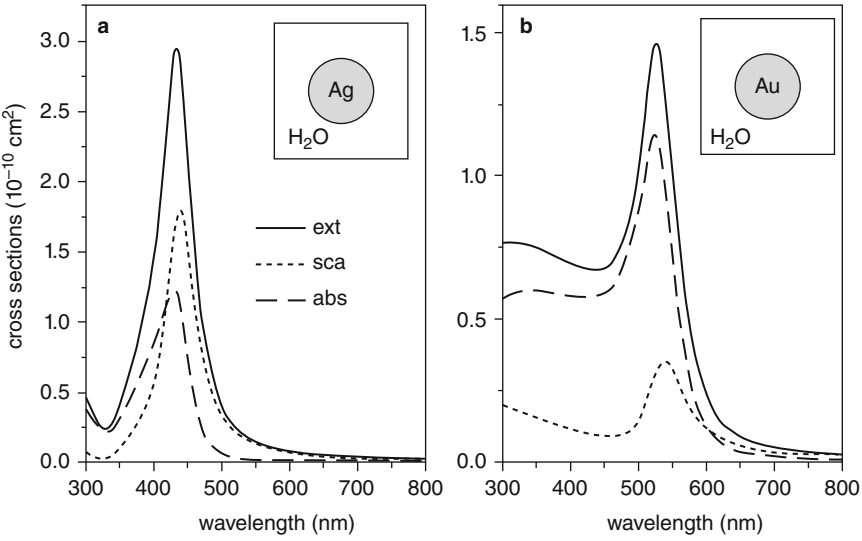


Fig. 5.20 Extinction, scattering, and absorption cross-sections of (a) a silver sphere of 60 nm diameter; (b) a gold sphere of 60 nm diameter. The two particles are in water. (From Klar[28])

where ω is the frequency of light, V is the volume of the particle, ε_{diel} is the (pure real) dielectric constant of the surrounding medium, and $\varepsilon_{met} = \varepsilon'_{met} + \varepsilon''_{met}$ is the complex dielectric constant of the metallic nanoparticle. The denominator in the case of the two relations (5.65a and b) can have a minimum in the case $\varepsilon'_{met} + 2\varepsilon_{diel} = 0$ because the real part of the refractive constant of noble metals is negative in the visible range. Figure 5.20 shows the extinction, scattering and absorption cross-sections for silver and gold spheres as a function of wavelength, respectively.

Inspection of the figure explains the colour properties of the Lycurgus cup illuminated by white light. The nanoparticles when illuminated from the behind absorb mainly green light and leave the reddish part of the spectrum to pass. When seen from the front side, they have a scattering maximum in the green and therefore appear principally of that colour.

Plasmon resonances of nanospheres with diameters smaller than 10 nm become more sensitive to the quantum nature of the conduction electrons.[29] A substantial blue shift in resonances is observed as particle size decreases, indicating the transition into the quantum influenced regime.

Plasmon interaction between near particles is an important subject not treated here.[30]

[28]Klar TA (2007) In: Shalev VM, Kawata S (eds) Nanophotonics with surface plasmons. Elsevier, Amsterdam

[29]Scholl JA et al (2012) Nature 483:421

[30]See for example, Zuloaga J et al (2009) Nano Lett. **9**:887, and references therein

5.9 Surface Phonon Polaritons in Dielectrics and Semiconductors

In the vibration spectrum of a solid one finds excitations of acoustic and optical type localized at surfaces. If we restrict ourselves to the long wavelength limit, corresponding to elastic vibrations of the continuum (acoustic branch) we find elastic surface waves, which propagate along the surface in a layer with the thickness of a wavelength. These are the so-called Rayleigh waves.

In the infinite medium we find two types of propagating waves, the longitudinal waves (irrotational component of the lattice vibrations, frequency limit ω_{LO}) and the transverse waves (divergence-free component, frequency limit $\omega_{TO} < \omega_{LO}$). Associated with them are the LO- and TO-phonons. Other types of excitations may also exist.

An electron in a crystal lattice, due to its electric charge interacts with the ions or atoms of the lattice and creates a local deformation of the lattice. This deformation tends to follow the electron which moves in the lattice. The combination of the electron and the associated deformation field is called *exciton polaron*.[31]

The interaction of the transverse electromagnetic field with the transverse polarization waves of the solid creates elementary excitations in the form of TO-phonons, for the polarization of the ion lattice, and the excitons, for the polarization of the electron system. Both are Bose particles. The interaction of the electromagnetic field then consists of photon-polarization quantum transformation and vice-versa.

The dispersion of polarization quanta is very weak and the dispersion curves of photons and polarization quanta intersect at low wave vector (Fig. 5.21).

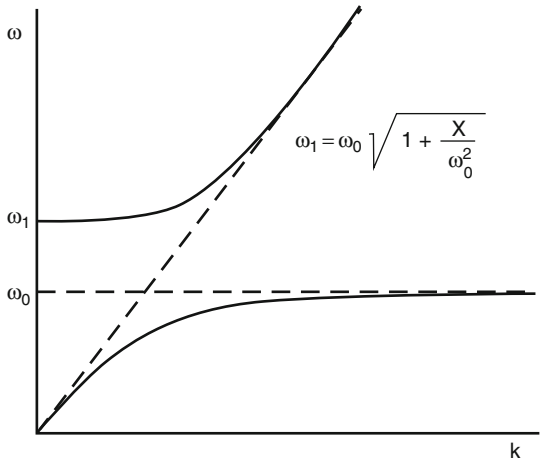


Fig. 5.21 Dispersion curves of polaritons in crystals

[31] See for example, Agranovich VM (1972) In: Abeles F (ed) Optical properties of solids. North-Holland Pu.Co, p 315; Agranovich VM, Mills DL (1982) Surface polaritons. North-Holland Pu.Co., Amsterdam

Near the point of intersection – where energy and momentum are the same for both excitations – the coupling becomes so strong that neither can continue to be regarded as an independent elementary excitation. Photon and polarization quantum rather exist together as a single quantity which can be interpreted as a new elementary excitation called *polariton*.

We describe the electromagnetic field by Maxwell equations and the polarization field by an equation of motion in which free oscillations of eigen-frequency ω_o are coupled with the electric field through a susceptibility χ

$$\nabla \times \mathbf{H} = \varepsilon_o \partial \mathbf{E}/\partial t + \partial \mathbf{P}/\partial t \tag{5.66a}$$

$$\nabla \times \mathbf{E} = -\mu_o \partial \mathbf{H}/\partial t \tag{5.66b}$$

$$\partial^2 \mathbf{P}/\partial t^2 + \omega_o^2 \mathbf{P} = \chi \varepsilon_o \mathbf{E}. \tag{5.66c}$$

We take the fields **E**, **H** and **P** to be transverse plane waves which propagate in the z-direction

$$E_x = E_{xo} \exp[i(kz - \omega t)] \tag{5.67a}$$

$$H_y = H_{yo} \exp[i(kz - \omega t)] \tag{5.67b}$$

$$P_x = P_{xo} \exp[i(kz - \omega t)]. \tag{5.67c}$$

Inserting (5.67) into (5.66) leads to

$$\omega \varepsilon_o E_x + \omega P_x - kH = 0 \tag{5.68a}$$

$$kE_x - \omega \mu_o H_y = 0 \tag{5.68b}$$

$$\chi \varepsilon_o E_x + (\omega^2 - \omega_o^2) P_x = 0. \tag{5.68c}$$

The system of equations only has a solution if the determinant

$$\begin{vmatrix} \omega \varepsilon_o & \omega & -k \\ k & 0 & -\omega \mu_o \\ \chi \varepsilon_o & \omega^2 - \omega_o^2 & 0 \end{vmatrix} \tag{5.69}$$

vanishes. Thus

$$\omega^4 - \omega^2(\omega_o^2 + \chi + c^2 k^2) + \omega_o^2 c^2 k^2 = 0 \tag{5.70}$$

from which the dispersion relation $\omega(k)$ for polaritons can be calculated. The dispersion is shown in Fig. 5.21. For large k the two branches of the dispersion spectrum

become identical to the branches $\omega = ck$ for photons, and $\omega = \omega_o$ for polarization quanta. For small k, on the other hand, the spectrum exhibits gross deviations. The one branch admittedly increases linearly from zero but its slope is not c but $c/[1 + (\chi/\omega_o^2)]^{1/2}$. The other branch starts at $\omega = \omega_o[1 + (\chi/\omega_o^2)]^{1/2}$.

If the elastic vibrations in the crystal are considered, one obtain the *phonon polaritons*. Highly doped semiconductors can exhibit plasma frequencies at mid- and far-infrared frequencies, albeit with high losses. Surface phonon polaritons exist which are coupled excitations of the electromagnetic field and phonon modes of polar materials such as SiC occurring at mid-infrared frequencies.

The use of semiconductors for low-frequency SPP propagation, apart from the enhanced confinement, is interesting because it offers the possibility to tune the carrier density and thus the plasma frequency by either thermal excitation, photo-carrier generation or directly carrier injection.

Here we derive the dispersion relation of surface modes for an isotropic crystal with a single dispersion oscillator.[32] It could be alkali halides, $A^{III}B^{V}$ semiconductors with zincblende structure, such as GaP, InSb, etc. The frequency dependent dielectric function in the form neglecting the damping and spatial dispersion is

$$\varepsilon(\omega) = \varepsilon_\infty + (\varepsilon_o - \varepsilon_\infty)\omega_{TO}^2/(\omega_{TO}^2 - \omega^2) \tag{5.71}$$

where ε_∞ and ε_o are the high frequency and static dielectric constants, respectively, and ω_{TO} is the frequency of the transverse $q = 0$ optical phonon.

The curve $\varepsilon(\omega)$ Fig. 5.22 has a pole at ω_{TO}. In the frequency range between ω_{TO} and ω_{LO} (longitudinal phonon frequency) $\varepsilon(\omega)$ is negative. It is this frequency range that is of interest.

Let us consider a semi-infinite crystal (ε_2) occupying the half-space $z < 0$. The plane $z = 0$ is the crystal-vacuum interface. Applying Maxwell's equations as we did in & Sect. 5.4 eventually we find the dispersion relation

$$\beta^2 = k_x^2 = (\omega/c)^2\varepsilon_1\varepsilon_2(\omega)/[\varepsilon_1 + \varepsilon_2(\omega)]. \tag{5.72}$$

It can be seen that the surface wave frequency for a given k value decreases with increasing ε_1. The dispersion dependence is schematically shown in Fig. 5.23.

The frequency range where the surface modes exist is limited by the roots of the equations $\varepsilon_2(\omega) = -\infty$ and $\varepsilon_2(\omega) = -\varepsilon_1$ (or $\varepsilon_2(\omega) = -1$ in case of crystal-vacuum interface). The lower limit coincides with the bulk transverse phonon frequency ω_{TO}. The upper limit is reached at $k > \omega_{TO}/c$ and will be designed as ω_s. Putting $\varepsilon(\omega)$ from (5.56) equal to $-\varepsilon_1$, we get for ω_s

$$\omega_s = [(\varepsilon_o + \varepsilon_1)/(\varepsilon_\infty + \varepsilon_1)]^{1/2}\,\omega_{TO}. \tag{5.73}$$

[32]Mirlin DN (1982) In: Agranovich VM, Mills DL (eds) Surface polaritons. North-Holland Pu.Co., Amsterdam, chapt. I

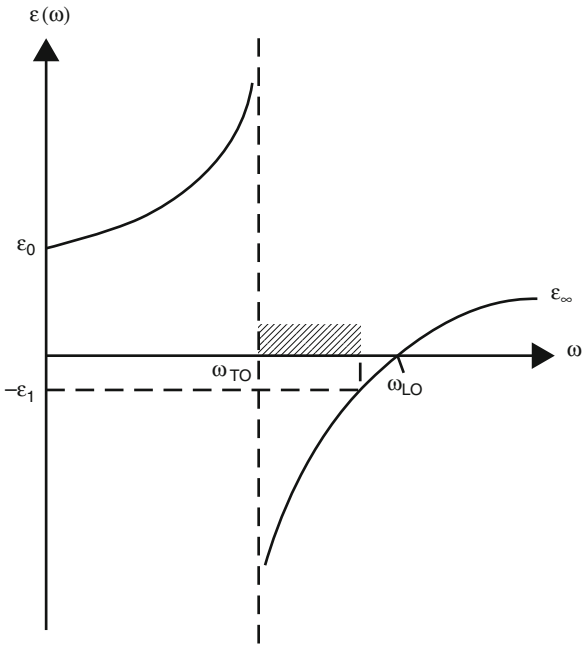


Fig. 5.22 Frequency dependence of the dielectric function of an isotropic crystal. The pole and zero of $\varepsilon(\omega)$ correspond to the transverse and longitudinal $q = 0$ optical vibrations respectively. The frequency range where surface modes exist is indicated by the shaded area

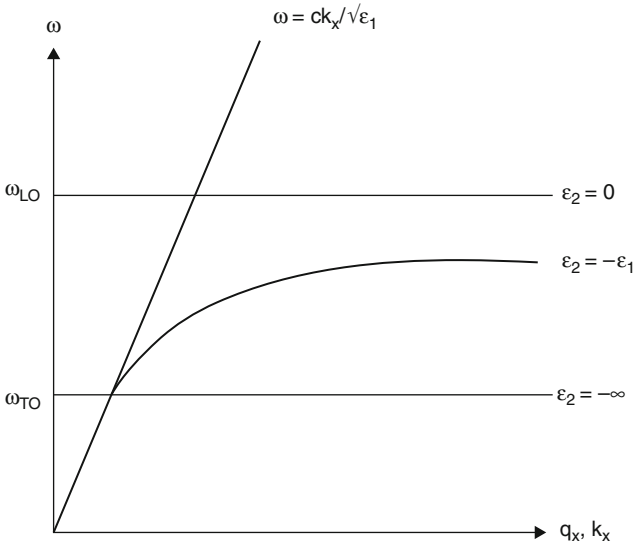


Fig. 5.23 The dispersion curve of the surface optical vibrations on a semi-infinite crystal bounded by a medium with dielectric constant ε_1. The position of $q = 0$ frequencies of bulk optical vibrations and the corresponding values of $\varepsilon_2(\omega)$ are shown as well as the light line $\omega = ck_z/(\varepsilon_1)^{1/2}$

The dispersion curve of the surface modes runs to the right of the light line $\omega = ck_x/\varepsilon^{1/2}$ (which corresponds to the light propagating along the surface of the crystal) and does not cross the light line anywhere. This fact implies that the vibration modes considered are not radiative, i.e. they cannot be radiated as photons and cannot couple with light. Coupling can be obtained using evanescent waves or gratings as described for metals.

In the case of a slab of finite thickness d, when d becomes comparable or smaller than $1/k_2$, two different branches of the surface modes with dispersion relations

$$\varepsilon_2(\omega)/\varepsilon_1 = -(k_2/k_1)\tanh(k_2 d/2) \tag{5.74a}$$

and

$$\varepsilon_2(\omega)/\varepsilon_1 = -(k_2/k_1)\coth(k_2 d/2) \tag{5.74b}$$

appear.

The frequencies of polaritons in crystals are usually in the far infrared. For example, in NaCl it is $\omega_{TO} = 164\ cm^{-1}$ and $\omega_{LO} = 255\ cm^{-1}$.

In the case of semiconductors in the long-wavelength limit and without damping we may write, taking into account the free-carrier contribution, the expression for the dielectric function as

$$\varepsilon(\omega) = \varepsilon_\infty + (\varepsilon_o - \varepsilon_\infty)\omega_{TO}^2/(\omega_{TO}^2 - \omega^2) - (\omega_p/\omega)^2 \tag{5.75}$$

the last term is due to the free-carrier contribution, $\omega_p = (4\pi ne^2/m^*)^{1/2}$ is the plasma frequency in vacuum, n and m* are the concentration and the effective mass of carriers, respectively. Putting $\varepsilon(\omega) = 0$ we obtain the equation for the bulk coupled plasmon-phonon modes. As long as the concentration of free carriers is low ($\omega_p < \omega_{LO}$), the frequency of the longitudinal lattice vibration obtained from $\varepsilon(\omega) = 0$ remains unperturbed. If ω_p and ω_{LO} are comparable, two mixed vibration branches occur.

Surface plasmon polaritons in the terahertz region can be obtained by doping semiconductors.[33] Doped semiconductors exhibit a behaviour at terahertz frequencies similar to that of metals at optical frequencies.

5.10 The Plasmons in Optical Nonlinear Materials

We consider here electromagnetic waves guided by a nonlinear interface.[34] Transverse electric (TE) polarized electromagnetic waves cannot be guided by the interface between two dielectric or metal/dielectric media in the linear case.

[33] Gomez-Rivas J et al (2004) Phys Rev Lett. 93:256804

[34] Mihalache D, Bertolotti M, Sibilia C (1989) Progress Optics 27:229

However, if one or both of the media exhibit an intensity-dependent refractive index, a TE polarized surface polariton may exist at powers exceeding a threshold value.

We show this here considering a nonlinear interface between an optically linear semi-infinite medium (called the substrate) with dielectric constant ε_s in region I ($z < 0$), a very thin metal film with dielectric constant $\varepsilon_f = -|\varepsilon_f| < 0$ in region II ($0 < z < d$), and a nonlinear self-focusing Kerr-law cladding in region II ($z > d$).

We remember that an optically nonlinear Kerr medium has an intensity dependent refractive index that can be written as

$$n = n_o + n_2 I \tag{5.76}$$

with

$$n_2 = 3\chi^{(3)}_{eff}/c\varepsilon_o n_o \tag{5.77}$$

and I is the light intensity

$$I = |E|^2/2(\mu/\varepsilon)^{1/2}, \tag{5.78}$$

$\chi^{(3)}_{eff}$ is the third-order nonlinear susceptibility, $n_2 > 0$ for self-focusing Kerr nonlinearity and $n_2 < 0$ for self-defocusing Kerr nonlinearity. Accordingly, the dielectric constant can be written as

$$\varepsilon_c = n_o^2 + \alpha|E|^2 \tag{5.79}$$

where

$$\alpha = c\varepsilon_o n_o^2 n_2. \tag{5.80}$$

The electric field in the three regions must fulfil the equations

$$\partial^2 E_y/\partial z^2 + (k_o^2\varepsilon_s - \beta^2)E_y = 0 \quad z < 0 \tag{5.81a}$$

$$\partial^2 E_y/\partial z^2 + (k_o^2\varepsilon_f - \beta^2)E_y = 0 \quad 0 < z < d \tag{5.81b}$$

$$\partial^2 E_y/\partial z^2 + (k_o^2\varepsilon_c - \beta^2)E_y + \alpha k_o^2 E_y^3 = 0 \quad z > d. \tag{5.81c}$$

In (5.81c) a term is added representing the nonlinearity.

The solutions satisfying boundary conditions are

$$E_y(z) = E_o \exp(k_o q_s z) \quad z < 0 \tag{5.82a}$$

$$E_y(z) = E_o[(q_s + q_r)\exp(k_o q_f z)/2q_r + (q_f - q_s)\exp(-k_o q_f z)/2q_f \quad 0 < z < d \tag{5.82b}$$

$$E_y(z) = (2/\alpha_c)^{1/2} q_c \{\cosh[k_o q_c (z - z_c)]\}^{-1} \quad z > d \tag{5.82c}$$

where

$$q_s = (\beta^2 - \varepsilon_s)^{1/2} \quad q_f = (\beta^2 + |\varepsilon_f|)^{1/2} \quad q_c = (\beta^2 - \varepsilon_c)^{1/2}. \tag{5.83}$$

Here the surface field E_o is given by

$$\alpha_c E_o^2 = 2\{1 - \tanh^2[k_o q_c (z_c - d)][\cosh(k_o q_f d) + (q_s/q_f)\sinh(k_o q_f d)]^{-2}. \tag{5.84}$$

As a result of the continuity requirements of E_y and dE_y/dz at the interfaces $z = 0$ and $z = d$, the effective index β is subject to a dispersion relation

$$\tanh(k_o q_f d) = q_f [q_c \tanh[k_o q_c (z_c - d)] - q_s]/[q_r^2 - q_s q_c \tanh[k_o q_c (z_c - d)]. \tag{5.85}$$

The dispersion relation has solution if, and only if, $n_s > n_c$ and $z_c > d$; that is a field maximum must always be situated within the cladding region.

The power per unit distance along the wavefront carried by the wave is given by

$$\begin{aligned} P_s &= (1/2) P_o \beta \alpha E_o^2 / q_s \\ P_f &= (1/2) P_o \beta \alpha_c E_o^2 \{k_o d(1 - q_s^2/q_f^2) + (1/q_f)\sinh(k_o q_f d)\big] \\ &\quad \times \big[(1 + q_s^2/q_f^2)\cosh(k_o q_f d) + 2(q_s/q_f)\sinh(k_o q_f d)\big]\} \\ P_c &= 2 P_o \beta q_c [1 + \tanh[k_o q_c (z_c - d)]] \end{aligned} \tag{5.86}$$

where

$$P_o = (2\alpha_c k_o)^{-1} (\varepsilon_o/\mu_o)^{1/2}. \tag{5.87}$$

The dependence of the propagation constant β on the guided wave power P is shown in Fig. 5.24.

A definite power threshold where the nonlinear TE polarized surface plasmon polariton starts is clearly seen. The threshold increases with refractive index of the substrate.

5.11 Other Surface Waves

Solving Helmoltz equation in the different media and applying the continuity conditions at boundaries, one may demonstrate that surface waves with decaying evanescent tails exist also at the interface between a great variety of media other than metal/dielectric. We just mention a few cases.

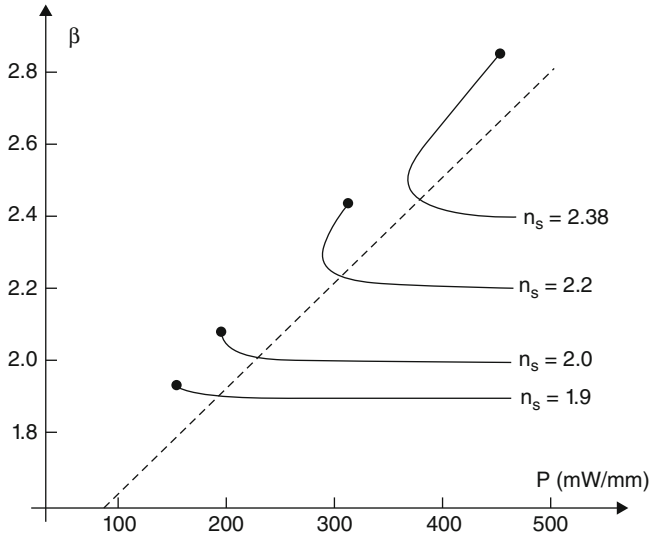


Fig. 5.24 Propagation constant β versus guided wave power for $n_c = 1.55$, $n_{2c} = 10^{-9}$ m^2/W, $\varepsilon_r = -10$, $\lambda = 515$ nm and $d = 0.005$ μm. Solid curves represent different n_s values and $n_{2s} = 0$. Dashed curve: $n_s = n_c = 1.55$ and $n_{2s} = n_{2c} = 10^{-9}$ m^2/W. (After Lederer and Mihalache[35])

5.11.1 Dyakonov Waves

A special type of surface wave was discovered by Dyakonov[36] in 1988. Such waves exist under special conditions at the interface of transparent anisotropic materials, and thus they are lossless.

Dyakonov[37] found that surface waves can form at interfaces between two dielectrics provided that at last one of the two media is an anisotropic medium, either a biaxial or positive uniaxial. In the case of the interface of a biaxial anisotropic medium with refractive indices n_x, n_y and n_z, and an isotropic medium with refractive index n_c, surface waves form only when the refractive indices fulfil the condition

$$n_1 > n_c > n_2 > n_3,$$

where $n_1 = \max(n_x, n_y, n_z)$, $n_3 = \min(n_x, n_y, n_z)$, and n_2 is between n_1 and n_3. For a positive uniaxial medium with indices n_o, n_e the following condition should instead be fulfilled

$$n_o < n_c < n_e. \tag{5.88}$$

[35]Lederer F, Mihalache D (1986) Solid State Comm. 59:151

[36]Dyakonov MI (1988) Sov Phys JETP 64:714

[37]Dyakonov MI (1988) Sov Phys JETP 64:714

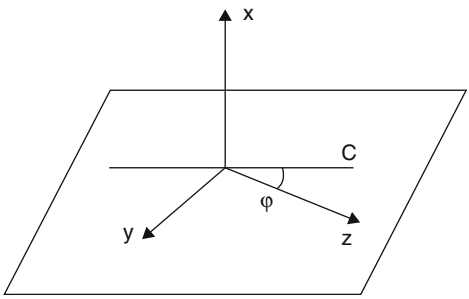


Fig. 5.25 Coordinate system. An isotropic medium is in the half-space x > 0 and a uniaxial crystal occupies the half-space x < 0. C is the optical axis of the crystal and the phase velocity of a surface wave is directed along z

Dyakonov waves feature unique properties: they are hybrid waves, i.e. they cannot be decoupled into either TE or TM modes, and they only exist for a well-defined range of orientations of the optical axis with respect to the propagation direction. However, the angular existence window is less than one degree for natural birefringent materials. This make the experimental excitation a challenge.

The existence of Dyakonov waves is due to the difference in symmetry of materials forming the interface. Later similar waves were predicted to exist at the interface between two identical uniaxial crystals with different orientations or a uniaxial material and a metal.[38]

Dyakonov surface waves also form in biaxial birefringent crystals.[39]

We shall consider the case of an interface between an isotropic medium with permittivity ε and a uniaxial crystal for which the principal values of the permittivity tensor are ε_e, ε_o and the optic axis is parallel to the interface x = 0 (Fig. 5.25). The half space x > 0 is occupied by the isotropic medium, whereas the crystal with its optic axis parallel to the interface is in the half-space x < 0. The direction of the phase velocity of a surface wave making an angle φ with the optic axis is assumed to be the z axis.

In the isotropic medium there are two independent solutions of the Maxwell equations with different polarizations and with the wave vector $q_1 = (ik_1, 0, q)$ where

$$q^2 - k_1^2 = \varepsilon(\omega/c)^2. \tag{5.89}$$

Two independent solutions in the crystal (extraordinary and ordinary wave) have the wave vectors $q_2 = (-ik_2, 0, q)$ and $q_3 = (-ik_3, 0, q)$ respectively, and are governed by the dispersion laws

$$(q^2\sin^2\varphi - k_2^2)/\varepsilon_e + (q^2\cos^2\varphi)/\varepsilon_o = 1 \tag{5.90a}$$

$$q^2 - k_3^2 = \varepsilon_o. \tag{5.90b}$$

[38]Liu HH, Chang HC (2013) IEEE Photonic J. 5:4800806

[39]Walker DB et al (1998) J Opt Soc America A15:248

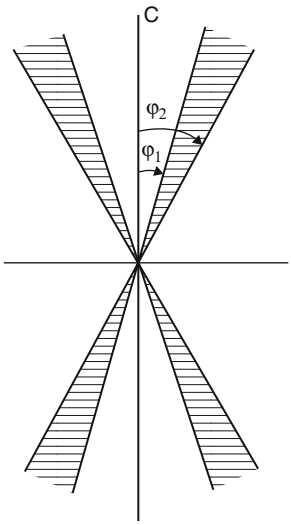


Fig. 5.26 Range of angles (shown shaded), relative to the optic axis where the phase velocity of a surface wave may be located

Finding the electric and magnetic field and applying the usual continuity conditions leads to

$$(k_1 + k_2)(k_1 + k_3)(\varepsilon k_3 + \varepsilon_o k_2) = (\varepsilon_e - \varepsilon)(\varepsilon - \varepsilon_o)k_3 \tag{5.91}$$

A discussion of this expression shows that solutions with k_1, k_2, $k_3 > 0$ corresponding to a wave localized at the interface exist if $\varepsilon_e > \varepsilon > \varepsilon_o$ or $\varepsilon_o > \varepsilon > \varepsilon_e$. It can be shown that only the first case gives solutions which, however can exist only in certain range of angles φ (Fig. 5.26).

Thus Dyakonov waves are supported by interfaces with positive birefringent crystals only ($\varepsilon_e > \varepsilon_o$). The angular interval is defined by the parameter $\eta = \varepsilon_e/\varepsilon_o - 1$.

In the case in discussion, one can show that if $\varphi = \varphi_1$, then surface waves transform into a plane wave in the isotropic medium, whereas for $\varphi = \varphi_2$, they transform into an extraordinary wave in the crystal.

The existence of Dyakonov waves has been proved by O Takayama et al.[40] Condition (5.86) does not commonly occur with natural materials, because the difference between the refractive indices of transparent anisotropic materials is typically small. Takayama used potassium titanyl phosphate (KTP) that is a biaxial crystal with relatively low refractive indices ($n_x = 1.7619$, $n_y = 1.7712$ and $n_z = 1.8648$ at 632.8 nm). A index-matching liquid with $n < 1.8$ was used as isotropic medium.

Figure 5.27 shows the experimental disposition. An index matching liquid is placed between a ZnSe prism ($n_p = 5.589$) and the KTP crystal. An expanded He-Ne beam ($\lambda = 632.8$ nm) is focused at the interface through a polarizer, a

[40]Takayama O et al (2009) Phys Rev Lett. 102:043903

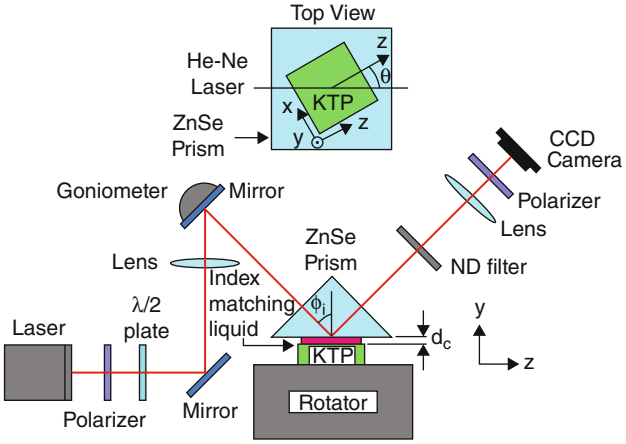


Fig. 5.27 The index matching liquid is placed between a ZnSe prism ($n_p = 2.589$) and the KTP crystal in a Otto-Kretchman configuration. A He-Ne beam is focused at the interface through a polarizer, a λ/2 plate, a lens and a mirror. The polarization of the reflected beam is analyzed. The crystal orientation θ (propagation angle) with respect to the incident beam is controlled with a rotator (da Takayama). (From O Tanaka et al., Phys. Rev. Lett. 102, 043903 (2009) reproduced with permission from AIP)[40]

λ/2 plate, a lens, and the ZnSe prism. The reflected beam is collected by an identical second lens, a polarizer, and directed to a CCD camera.

Surface wave excitation was detected by imaging the polarization-conversion reflectance spectrum. Because of the biaxial anisotropy of the substrate, the radiation to the prism associated to the half-leaky modes results in a polarization conversion from TE or TM incident light to TM or TE reflected light, respectively.

5.11.2 Negative Index Materials

Depending on the material parameters, TE or TM surface waves can form at the interface between left handed materials and dielectrics,[41] although not simultaneously (i.e. not for the same set of parameters). These particular surface waves have been studied by several people.[42] The energy fluxes in the dielectric and NIM half-spaces are in opposite directions at the NIM and regular dielectric interfaces. This vortex like behaviour is another unique feature of an interface NIM/dielectric interface. Moreover Shadrivov et al.,[43] have shown by direct numerical

[41]Ruppin R (2000) Phys Lett. A277:61

[42]Lindell IV et al (2001) Microwave Opt Tech Lett. 31:129; Shadrivov IV et al (2003) Phys Rev. E67:057602; Darmanyan S et al (2003) Opt Comm. 252:233; Shadrivov IV et al (2004) Phys Rev. E69:016617; Shadrivov IV et al (2005) Opt Expr. 13:481; Bordman A et al (2005) New J Phys. 7:191

[43]Shadrivov IV et al (2005) Opt Express 13:481

simulations that, under appropriate conditions, a TE surface wave can be excited from an incident Gaussian beam. Surface waves at interfaces between left-handed materials and dielectric birefringent media may exist with existence conditions highly relaxed in comparison to Dyakonov waves in configurations involving birefringent materials and right-handed media. The existence of several hybrid surface waves was predicted by Crasovan et al.[44]

Surface plasmon polaritons in the terahertz region can be obtained by doping semiconductors[45] or putting high-index dielectric overlayers on flat metallic films. A thin metal sheet perforated with an array of holes leads to the appearance of confined surface modes,[46] versatile plasmonic metamaterials can be built consisting of metal surfaces textured with subwavelength-scale corrugation or dimples that have surface waves that mimic the properties of SPPs. An example is given by Williams et al.[47] who have studied SPP propagation on a planar copper surface perforated by square arrays of square lined holes demonstrating propagation and confinement of terahertz electromagnetic surface modes with an effective plasma frequency controlled entirely by the surface geometry.

5.11.3 Tamm Plasmon Polaritons

Surface waves can be produced at the interface between a planar multilayer structure forming a dielectric Bragg reflector and a metal[48] and are called *Tamm plasmon polaritons* (TPP) from their analogy with the electronic Tamm states localized at crystal surfaces. In contrast to conventional surface plasmon, TPPs can be formed in both the TE and TM polarizations and have an in-plane wavevector less than the wavevector of light in vacuum, which allows for their direct optical excitation.

More generally they can be formed at the interface between two periodic structures. The first has a period close to the wavelength. The second has a period close to the double of the wavelength. Cladding and substrate must have overlapping band gaps, but different periods. They exist in any direction along surface. They are highly sensitive to the order of the layers at the interface. The dispersion curve is parabolic.[49]

They have been observed by Sasin[50] at the interface between a metal and a dielectric Bragg reflector.

[44]Crasovan LC et al (2005) Forse Opt Express, CERCARE
[45]Gomez-Rivas J et al (2004) Phys Rev Lett. 93:256804
[46]Ulrich R, and Tacke M (1953) Appl Phys Lett. 22:251
[47]Williams CR et al (2008) Nature Phot. 2:179
[48]Kaliteevski M et al (2007) Phys Rev. B76:165415
[49]Kavokin AV et al (2005) Phys Rev. B72:233102
[50]Sasin ME et al (2008) Appl Phys Let. 92:251112

5.11.4 Plasmons in Graphene

The first isolation in 2004 of individual flakes of graphene, a single atomic layer of carbon atoms, spurred a great interest in studying its properties. By using graphene as a flatland platform, one can have propagating SPPs on a surface of the thickness of a carbon atom. To represent the dynamics of electrons in this material one commonly uses its complex conductivity $\sigma_g = \sigma_{gr} + i\sigma_{gi}$. The dispersion relation for the highly confined TM SPP waves along a free standing graphene in free space has been derived as[51]

$$\beta^2 = k_o^2\left[1 - (2/\eta_o\sigma_g)^2\right]$$

where k_o and η_o are respectively the free space wave number and the intrinsic impedance of free space.

The advantages of SPPs on graphene over plasmons in conventional materials include relative low losses,[52] strong light-matter interactions,[53] electronic tunability,[54] and large field confinement at THz and infrared frequencies. Coupling and propagation of the fundamental plasmonic mode along a graphene strip of finite width transferred onto a dielectric has been discussed by J S Gomez-Diaz et al.[55]

It also possible to excite surface waves in many other materials using liquid crystal,[56] uniaxiaslly anisotropic,[57] gyrotropic,[58] chiral,[59] etc. interfaces.

[51]Hanson GW (2008) J Appl Phys. 103:064302; Vakil A, Engheta N (2011) Science 332:1291; Mikhailov SA, Ziegler K (2007) Phys Rev Lett. 99:016803; Jablan M et al (2009) Phys Rev. B80:345435

[52]Christensen J et al (2013) ACS Nano 6:431

[53]Koppens FH et al (2011) Nano Lett. 11:3370

[54]Novoselov KS et al (2004) Science 306:666

[55]Gomez-Diaz JS et al (2013) Opt Express 21:24856

[56]Faryad M, Lakhtakia A (2011) Phys Rev. A83:013814

[57]Liu HH (2013) IEEE Phot J. 5:4800806

[58]Boardman A et al (2005) New J Phys. 7:191

[59]Engheta N, Pelet P (1991) Opt Lett. 16:723

Chapter 6
Applications of Plasmons

6.1 Introduction

Plasmons have a number of applications in many directions. Physicists, chemists, material scientists and biologists all are considering them for different purposes. One group of applications is in plasmon circuitry. It means transmission and operations on SPPs mainly for telecommunications and computers. A second group of applications uses SPPs as a mean to obtain new effects or to improve existing operations. Examples are surface enhanced Raman scattering and enhanced transmission through nanoscale apertures. In a third group we may include all the applications which use localised surface plasmons. For example localized surface plasmons can be used for enhancing the diffraction efficiency in holograms,[1] or to use the colour selectivity[2] of nanoparticles for holographic colour reconstruction with white light illumination. Plasmon holography in this case, does not suffer from the ghost produced by the diffraction of ambient light in higher orders of diffraction, because those components are not coupled with SPPs.

We will here discuss only some examples; namely surface enhanced Raman scattering (SERS), sensors, enhanced transmission through nanoscale apertures, plasmon circuitry, lasers, application to solar cells, plasmon microscopy, and some properties of coherence and radiation management. Many other applications

[1]Cowan JJ (1972) Opt Comm. 5:69; Maruo S et al (1997) Appl Opt. 36:2343; Wang GP et al (2001) Appl Opt. 40:3649

[2]Ozaki M et al (2011) Science 332:218

© Springer International Publishing AG 2017

M. Bertolotti et al., *Evanescent Waves in Optics*, Springer Series in Optical Sciences,
DOI 10.1007/978-3-319-61261-4_6

exist like optical tweezers,[3] nanolithography,[4] data storage,[5] nanoantennas,[6] optical spectrum analyser,[7] just to quote some.

6.2 Surface Enhanced Raman Scattering (SERS)

Surface enhanced Raman scattering (SERS) is probably the most spectacular effect in which plasmons are involved. In 1928, the Indian physicist C V Raman[8] discovered that a very tiny fraction of light scattered by molecules, contained frequencies that where increased or decreased with respect to the frequency of the incoming light by a small amount corresponding to one of the vibrational frequencies of the molecule. The explanation of the effect is very simple: the incoming photon may lose energy exciting a vibration level of the molecule and therefore its frequency is decreased (*Stokes line*) or it may gain energy from the vibration, increasing its frequency (*anti-Stokes line*) (Fig. 6.1). The phenomenon, called *Raman effect*, provides a rich variety of information on the structure and composition of matter, gained from the knowledge of its vibration spectrum that is so measured. However typical Raman cross-sections per molecule, range between

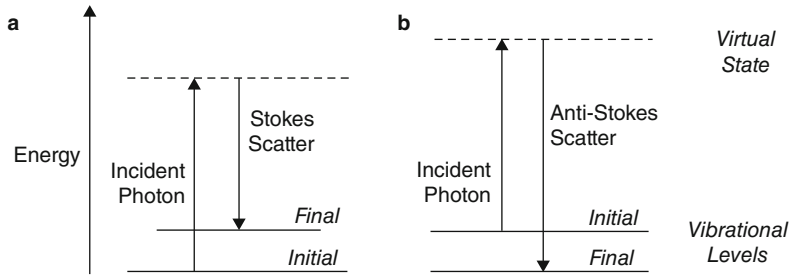


Fig. 6.1 Energy level diagram for Raman scattering. (**a**) Stokes Raman scattering. The incident photon is absorbed in a transition to an upper level. In the decay process it ends on an excited vibrational level. Conservation of energy reads $h\nu = h\nu' + E_{ex}$, where ν and ν' are the frequencies of the incoming and scattered photon, respectively, and E_{ex} is the energy of the excited vibrational level. (**b**) Anti-Stokes Raman scattering. In this case the incident photon is absorbed in a transition from an excited vibrational state to an upper state, and the scattered photon ends on the ground state. Conservation of energy is $h\nu + E_{ex} = h\nu'$

[3]Reece PJ (2008) Nat Photon. 2:333; Juan ML et al (2011) Nat Photon. 5:349

[4]Ozbay W (2006) Science 311:189; Ni X et al (2012) Science 335:427

[5]Zijlstra P et al (2009) Nature 459:410

[6]Toussaint Jr KC et al (June 2015) Optics and photonics news. p. 24

[7]Li JJ et al (2009) Phys Rev. B79:195104

[8]Raman CV, Krishnan KS (1928) Nature 121:501

10^{-30} and 10^{-25} cm^2, with the larger values occurring when the frequency of light happens to match an electronic transition in the molecule (*resonant condition*). The use of the laser dramatically improved the signals which, however remained extremely small.

A turn occurred in 1974 when Fleishmann et al.,[9] observing the Raman spectra of pyridine, reported a huge increase in the scattering cross-section if the molecules are adsorbed on roughened metal surfaces, and explained this finding assuming the useful surface area was increased by the roughness. Van Duyne[10] and independently Creighton[11] were the first to recognize that a new effect was present, demonstrating that the increase of useful area was not sufficient to explain the experimental results. The phenomenon received the name of SERS (Surface Enhanced Raman Scattering). Typical enhancement factors for the Raman signal observed from rough metal substrates as compared to bare glass substrates are of the order of 10^6–10^7, and using resonance enhancement (excitation frequency near an electronic transition frequency) enhancement factors as high as 10^{12} have been reported. Using chemically roughened silver surfaces, Raman scattering events of single molecules have been recorded,[12] with estimated enhancements of the scattering cross-section by factors up to 10^{14}.

The effect received considerable attention and decades of debate on its origin.

The phenomenon is now mostly understood: the majority of the enhancement has been realized to arise from the highly enhanced fields near metal nanoparticles due to localized surface plasmon resonances. Termed, *hot spot*, these highly confined fields also enable an increase of fluorescent emission, albeit with more modest enhancement factors.

SERS is today a very powerful spectroscopic method that can be used even at single molecule level and is applied to ultrasensitive detection.[13]

If we limit the discussion to the power scattered in the Stokes process P_S (Raman line of frequency lower than the pumping frequency) we may write

$$P_S(\nu_S) = N\sigma_{RS} I(\nu_L), \tag{6.1}$$

where N is the number of Stokes-active scatterers within the excitation spot, σ_{RS} is the scattering cross-section, and $I(\nu_L)$ is the intensity of the excitation beam.

SERS describes the enhancement of this process, accomplished by placing the Raman-active molecules within the near-field of a metallic nanostructure.

[9]Fleischmann M, Hendra PJ, McQuillan AJ (1974) Chem Phys Lett. 26:163

[10]Jeanmaire DL, Van Duyne RP (1977) J Electroanal. Chem 84: 1

[11]Albrecht MG, Creighton JA (1977) J Am Chem Soc. 99: 5215

[12]Kneipp K et al (1996) Phys Rev Lett 76:2444; Kneipp K et al (1997) Phys Rev Lett. 78:1667; Nie SM, Emory SR (1997) Science 275:5303; Nie SM, Emory SR (1997) Science 275:1102; Kneipp K et al (1998) Phys Rev. E57:R6281; Michaels AM et al (1999) J Am Chem Soc. 121:9932 (1999); Xu HX et al (1999) Phys Rev Lett. 83:4357

[13]Fischer (1986) J Opt Soc America B3:1239; Levene et al (2003) Science 299:682

The nanostructure can consist of metal colloids, specifically designed nanoparticle ensembles, or topography of a roughened surface. The enhancement of P_S is due to two main causes: a modification of the Raman cross-section due to a change in environment of the molecule, and, more important, an increased electromagnetic field due to the excitation of localized surface plasmons and a crowding of the electric field lines at the metal interface.[14] This leads to an enhancement of both the incoming and emitted light fields. Since the frequency difference between the incoming and scattered photons is in general much smaller than the line-width of a localized surface plasmon mode, the electromagnetic contribution to the total SERS enhancement is proportional to the fourth power of the field enhancement factor.

In fact, the field of the incident light, because of nanoplasmonic resonances, is enhanced in the close vicinity of the nanoparticle. This will cause an enhancement of the Raman effect because the Raman signal depends linearly on the intensity of the incoming light, which is, in the case of nearby metallic structures, proportional to the square of the enhanced near field. Then the molecule will eventually alter the frequency of a photon by adding or subtracting a vibronic quantum and create a photon with the Raman frequency. In the emission of the Raman photon the nanoparticle will act as an antenna to increase the Raman field and efficiently radiate it. The efficiency of this radiative effect is again proportional to the local field intensity, this time at the Raman frequency. Therefore the total enhancement of the Raman signal of a molecule close to a metal particles in proportional to the enhancement of the illumination field squared times the enhancement of the Raman field squared, which explain why the total enhancement is the fourth power of the enhancement of the incoming field.

However, in general, the optimum field enhancement for the illumination field and the Raman field may not spatially coincide and one has to take the spatial average over the active volume. This decreases the overall enhancement but shows that the local enhancement is much higher of what is actually measured.

The enhancement depends on the type of metal, its degree of roughness, the size and shapes of the clusters and the frequency of the incident light which should be in the plasmon resonance of nanometer-sized bumps on the rough surface.

The subject is widely treated in many places.[15]

[14]Kerker et al (1980) Appl Opt. 19:4159; Gersten J, Nitzan A (1980) J Chem Phys. 73:3023; Weitz DA et al (1983) J Chem Phys. 78:5324

[15]Kneipp K et al (2002) J Phys Cond Matt. 14:R597; Kneipp K (Nov. 2007) Physics Today, p. 40; Schlucker S (2014) Ang Chem Int Edition 53:4756; Lee J et al (2011) Chem Eur J. 17:2381; Sharma B et al (2012) Materials Today 15:16; see also Moskovits M (1985) Rev Mod Phys. 57:783, A few books are Aroca R (2006), Surface-enhanced vibrational spectroscopy. Wiley, New York; Surface-enhanced Raman scattering physics and applications, vol. 103. In: Kneipp K, Moskovits M, Kneipp H (eds) Topics in applied physics. Springer, Berlin 2006; Le Ru E, Etchegoin P (2009) Principles of surface enhanced Raman spectroscopy and related plasmonic effects. Amsterdam

6.3 Surface Plasmon Sensors

Optical sensors based on surface plasmons, often referred to as *surface plasmon resonance (SPR) sensors*, exploit the sensitivity of the propagation constant of a surface plasmon to refractive index changes occurring in its evanescent field. A change in the refractive index produces a change in the propagation constant of the surface plasmon, which results in a change in the characteristics of the light wave interacting with it. The first sensor of this kind was demonstrated by Nylander and Liedberg[16] for gas detection.

The simplest scheme of the optical system consists of a light source, an SPR coupler (prism or grating), and a detector. The light source provides a beam of light which is introduced to the SPR coupler to excite a surface plasmon. Subsequently the reflected or diffracted light is detected and analyzed by the detector.

It is possible to measure many different properties of the light wave interacting with the surface plasmon, such as its angular behaviour, wavelength, intensity, phase and polarization modulations.

For example, the sharp resonance conditions associated with the excitation of surface plasmons can be used to make sensors in which the angular behaviour of the interacting light is the measured parameter. The position of the dip in the reflectivity curves as the incidence angle is varied can be used as an indicator of the environment changes. Figure 6.2 shows that a small change in the refractive index at the surface of the metal film, affecting the propagation constant of the surface plasmon, produces a quite measurable change in the resonant angle, and a shift of the angular position of the dip.

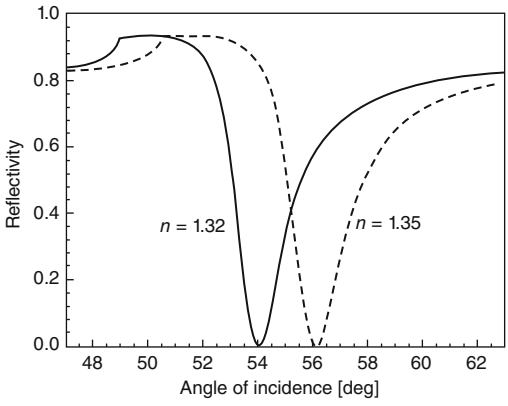


Fig. 6.2 Reflectivity in the prism coupling geometry with a glass prism on a 50 nm thick gold layer and a dielectric, as a function of the angle of incidence for two different refractive indices of the dielectric. Wavelength 632 nm

[16]Nylander C, Liedberg B, Lind T (1983) Sensors Actuators 3:79

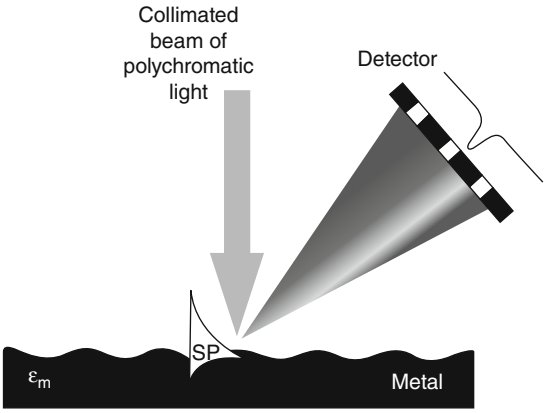


Fig. 6.3 Scheme of a complex sensor based on simultaneous excitation via second order diffraction on a special metallic grating of surface plasmons by a polychromatic light and simultaneous observation of the first order diffracted light. The coupling into a SPP results in a drop of the intensity of diffracted light. A refractive index resolution of 3.10^{-7} RIU was obtained. (From Telezhnikova and J Homola, J. Opt. Lett. 31, 3339 (2006) © Optical Society America)

With this method, the adsorption or removal of target materials on the metal surface can be detected with sub-monolayer accuracy.

The change in coupling conditions can also be seen as a change in the coupling wavelength (Fig. 6.3), or in the intensity or phase of the coupled light.

The main performance characteristics of sensors include the sensibility, linearity, resolution, accuracy, reproducibility, dynamic range and limit of detection.

The sensor *sensibility* is the ratio of the change in sensor output to the change in the quantity to be measured (e.g. the refractive index). The sensitivity of an SPR sensor to a refractive index S_{RI} can be written as

$$S_{RI} = (\delta Y/\delta n_{eff})(\delta n_{eff}/\delta n) \tag{6.2}$$

where Y is the sensor output and n_{eff} is the propagation index of the plasmon. The first term describes the sensitivity of the sensor output to the effective index of the surface plasmon and depends on which method of excitation of the surface plasmon is used. The second term describes the sensitivity of the effective index of a surface plasmon to the probed refractive index.

The *resolution* is defined as the smallest change in the refractive index that produces a detectable change in the sensor output. This is linked to the limit of detection. The magnitude of the sensor output change that can be detected depends on the level of uncertainty of the sensor output: the output noise.

Currently, the best SPR sensors using conventional surface plasmons achieve a performance level which, when expressed in terms of the smallest detectable refractive index change (refractive index resolution) approaches 10^{-7} refractive index units (RIU).[17]

[17]Homola J (2008) Chem Rev. 108:462

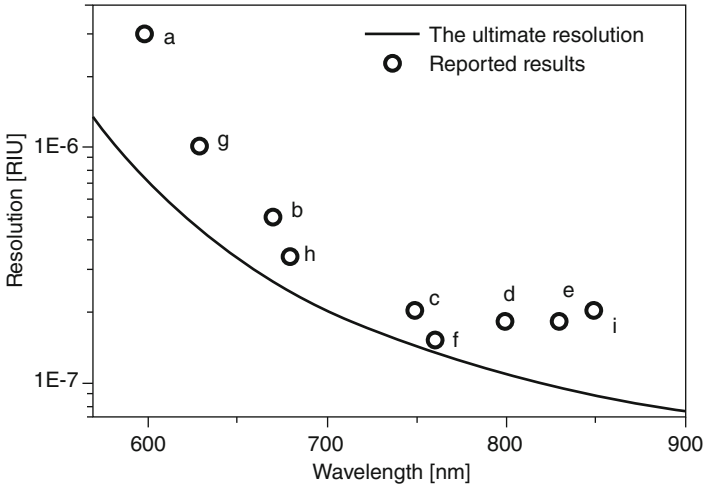


Fig. 6.4 The resolution obtained in some experimental results (**a**) Stemmler et al.,[18] (**b**) Thirstrup et al.,[19] (**c**) Piliarik et al.,[20] (**d**) Nenninger et al.,[21] (**e**) Chinowsky et al.,[22] (**f**) Biacore,[23] (**g**) Wu et al.,[24] (**h**) Bardin et al.,[25] (**i**) Piliarik et al.[26] (From M Piliarik and J Homola, Opt. Express 17, 16505 (2009) © Optical Society America)

Figure 6.4 shows the resolution obtained by several researchers as a function of wavelength. The continuous line represents the ultimate resolution that can be achieved as calculated by Piliarik and Homola.[27]

Coupling can be made by using a prism, a waveguide or a grating. Prism couplers represent the most frequently used method for optical excitation of surface plasmons.

The coupling of fluorophores to the surface can drastically improve the sensitivity of fluorescence based arrays in medical diagnostics, biotechnology and gene expressions. For finite distances between metal and fluorophores (<200 nm) the coupling to surface plasmons leads to fluorescence signal enhancement and high directionality of the emission.

[18]Stemmler I et al (1999) Sensors Actuators B Chem. 54:98

[19]Thirstrup C et al (2005) Sensors Actuators B Chem. 106:796

[20]Piliarik M et al (2009) Biosens Bioelectron. 24:1399

[21]Nenninger GG et al (2002) Meas Sci Technol. 13:2038

[22]Chinowsky TM et al (2003) Sensors Actuators B Chem. 91 :266

[23]Biacore 3000 (GE Healthcare, USA) has been developed by Swedish BIAcore AB (originally Pharmacia Biosensor AB) which was the first firm to launch a commercial SPR biosensor on the market in 1990.

[24]Wu CM et al (2004) Opt Express 12:3509

[25]Bardin F et al (2009) Biosens Bioelectron. 24:2100

[26]Piliarik M et al (2009) Biosens. Bioelectron. 24:3430

[27]Piliarik M, Homola J (2009) Opt Express 17:16505

Surface plasmon resonance sensors are today commercially available by several firms.[28]

Surface plasmon resonance sensors can be constructed using also fibres[29] or integrated optical waveguides[30]. Fiber optic SPR sensors present the highest degree of miniaturization. They also allow sensing in inaccessible locations where the ability to transmit optical signals over a long distance makes the use of optical fibres very attractive. The use of optical fibres for SPR sensing was first proposed by Jorgenson and Yee.[31] In SPR sensors based on side-polished single-mode optical fibres, a fundamental mode of the fibre couples to a surface plasmon at the outer surface of a metal layer deposited on a side-polished region of the fibre. The coupling results in attenuation of a transmitted light at a fixed wavelength (intensity-modulated SPR sensor) or a characteristic dip in the spectrum of transmitted light (wavelength-modulated SPR sensor). The main challenge for obtaining a stable performance is the sensitivity of polarization of light guided in the fibre to deformations of the fibre.

Plasmon sensors are used for bio-recognition, food quality and safety analysis, medical diagnostic, environmental monitoring, detection of chemical and biological species. They are described in a vast literature.[32]

A few specific examples[33] may help to understand these applications. A DNA microarray based system has been made using plasmon resonant particles. The optical observation of individual particles constitutes the detected signal using a CCD camera. The software measures the colour, brightness, size and location of all scattering entities.[34]

Gold nanoparticles are commonly used as signal-amplification labels in conventional propagating SPR. They are also employed in localized SPR. The peaks of gold nanoparticles are strongly governed by their size, shape and the dielectric constant of the surrounding medium. One can therefore find the right size of gold nanoparticles and couple their resonance to the surface plasmon waves for obtaining the largest field enhancement effect thereby increasing the sensitivity of the sensor.[35]

[28]Homola J (2008) Chem Rev. 108:462

[29]Jorgenson RC, Yee SS (1993) Sensors Actuators B12:213; Slavik R, Homola J, Ctyroky J (2001) Sensors Actuators B74:106

[30]Dostalek J et al (2001) Sensors Actuators 76:8

[31]Jorgenson RC, Yee SS (1993) Sensors Actuators B12:213

[32]Liedberg B et al (1983) Sensors Actuators 4:299; Homola J et al (1999) Sensors Actuators B54:3; Homola J, Yee SS, Myszka D (2002). In: Ligler FS, Taitt CR (eds) Optical biosensors: present and future. Elsevier; Homola J (2006) Surface plasmon resonance based sensors. Springer; Zeng S et al (2014) Chem Soc Rev. 43:3426

[33]Matveeva E et al (2004) Angew Chem. 76:6287

[34]Oldenburg SJ et al (2002) Analyt Biochem. 309:109

[35]Zeng S et al (2014) Chem Soc Rev. 43:3426

6.4 Extraordinary Optical Transmission Through Arrays of Sub-wavelength Holes

The transmission of light through a single hole (aperture) in an opaque screen is determined by diffraction. If the hole has a radius r much larger than the wavelength, for normally incident monochromatic plane-wave light, the transmitted intensity per unit solid angle in the far field is given by the well known formula

$$I(\theta) = I_o(k^2r^2/4\pi)|2J_1(krsin\,\theta)/krsin\,\theta|^2 \tag{6.3}$$

where I_o is the total intensity impinging on the aperture of area πr^2, θ is the angle between the aperture normal and the direction of the re-emitted radiation and $J_1(krsin\,\theta)$ is the Bessel function of the first kind. The function described by (6.3) gives place to the well known Airy pattern. The ratio of the total transmitted intensity to I_o, given by

$$T = (1/I_o)\int I(\theta)d\Omega \tag{6.4}$$

is called the *transmission coefficient*. For apertures with $r \gg \lambda$, in which case this treatment is valid, $T \sim 1$. For holes of $r < \lambda$, already in 1944, Hans Bethe[36] discussed the problem and found that assuming I_o constant over the area of the aperture, for normal incidence the aperture can be described as a small magnetic dipole located in the plane of the hole. The transmission coefficient for an incident plane wave is

$$T = (64/27\pi^2)(kr)^4 \sim (r/\lambda)^4. \tag{6.5}$$

The scaling with $(r/\lambda)^4$ implies very weak total transmission, smaller by a factor $(r/\lambda)^2$ compared to Kirchhoff theory for a sub-wavelength aperture. Equation (6.5) is valid for both TE and TM polarization.

Bethe calculation assumed the thickness of the conducting screen to be infinitely small. Taking into account the finite thickness h of the screen and the skin depth δ of the metal, the transmission coefficient T decreases exponentially with h[37]

$$T = \exp(-h/\delta). \tag{6.6}$$

A behaviour characteristic of a tunnelling process. If the diameter d of the hole allows propagating modes inside it (waveguide aperture), calculation shows that

[36]Bethe HA (1944) Phys Rev. 66:163. See also Bouwkhamp CJ (1950 Philips Recs Dep. 5321:401)

[37]Roberts A (1987) J Opt Soc America 4:1970

the transmission is increased. For optically thick films of a real metal with h of the order of several skin depths, radiation tunnelling is prevented and localized surface plasmons significantly influence the transmission process.[38]

In 1998, Ebbesen et al.[39] discovered an anomalous high light transmittance through a periodic array of holes with a size smaller than the diffraction limit. After a hot debate, a general consensus has been reached that the effect may be explained invoking surface plasmons even if the exact modalities are still under discussion.[40]

For light of wavelength λ, a hole less than λ/2 in diameter in a thick metal film does not support any propagating modes; energy can propagate only by an evanescent tunnelling process leading to very weak transmittance, in marked contrast to the experimental observations. The enhanced transmission should therefore be accounted for, by invoking the role of SPPs.

The transmission of light through a sub-wavelength aperture of a geometry such as a circle or a square that does not allow a propagating mode can be dramatically enhanced by structuring the screen with a regular, periodic lattice. This way, SPPs can be excited due to the grating coupling, leading to an enhanced light field on top of the aperture. After tunnelling through the aperture, the energy in the SPP field is scattered into the far field on the other side.

The phase-matching condition imposed by the grating leads to a well-defined structuring of the transmission spectrum T(λo) of the system, with peaks at the wavelengths where excitation of SPP takes place. At these wavelengths, $T > 1$ is possible, more light can tunnel though the aperture than incident on its area, since light impinging on the metal screen is channelled through the aperture via SPPs. This extraordinary transmission property was first demonstrated by Ebbesen et al. for a square array of circular apertures in a thin silver screen.[41]

Figure 6.5a shows the transmission spectrum for normally incident light on a silver screen of thickness $h = 200$ nm perforated with an array of circular holes of diameter $d = 150$ nm arranged on a square lattice with period $a_o = 900$ nm (Fig. 6.5b).[42]

In Fig. 6.5a the peak on the left at 326 nm is the narrow bulk silver plasmon peak and disappears as the film becomes thicker. At longer wavelengths, the spectrum is characterized by a set of maxima and minima of which the positions are determined by the geometry of the array. The peaks are at the wavelengths that correspond to phase matching of the incident light with surface plasmons on both

[38]Degiron A et al (2004) Opt Comm. 239:61; Degiron A et al (2002) Appl Phys Lett. 81:4327

[39]Ebbesen TW et al (1998) Nature 391:667

[40]Barnes WL et al (2003) Nature 424:824; Ebbesen TW et al (May 2008) Phys Today, p. 44; Mary A et al (2007) Phys Rev. B76:195414; Sturman B et al (2008) Phys Rev. B77:075106; Stockman MI (Feb 2011) Phys Today. p. 39; Maier SA (2007) Plasmonics: fundamentals and applications. Springer

[41]Ebbesen TW et al (1998) Nature 931:667

[42]Ghaemi HF et al (1998) Phys Rev. B58: 6779

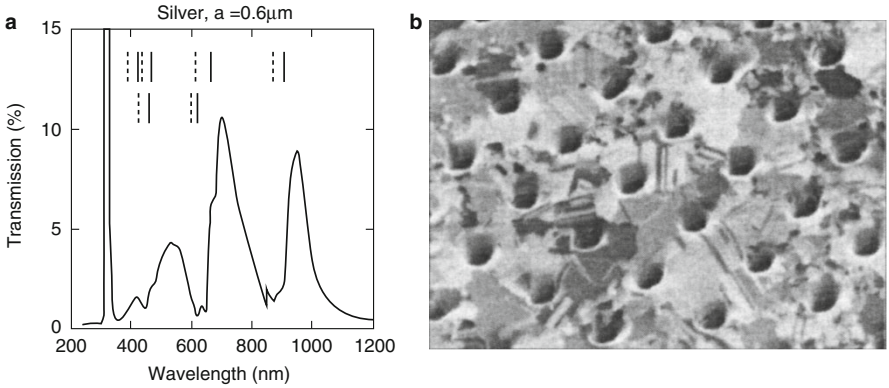


Fig. 6.5 Transmission of a two-dimensional hole array in a polycrystalline Ag film at normal incidence (**a**) and focused ion beam image of the Ag film. (From H F Gaemi et al., Phys. Rev. B58, 6779 (1998) reproduced with permission from AIP)

surfaces of the metal film with the help of the grating, The surface plasmon wave-vector β obeys the relation

$$\beta = k_x \pm nGx \pm mGy = k_o \sin\theta \pm (n+m)2\pi/a_o \tag{6.7}$$

where θ is the incidence angle of the radiation, $k_x = k_o \sin\theta$ is the field component of the incident photon wave-vector in the plane of the grating and $G_x = G_y = 2\pi/a_o$ are the grating momentum wave-vectors. Exciting SPPs allows light impinging on opaque regions between the apertures to be channelled to the other side via propagating SPPs, so explaining the increased transmission. Using the surface plasmon dispersion given in Sect. 5.4 at normal incidence θ = 0 one has for the wavelengths of the maxima

$$(n^2+m^2)^{½}\,\lambda_{max} = a_o[\varepsilon_1\varepsilon_2/(\varepsilon_1+\varepsilon_2)]^{½} \tag{6.8}$$

which with the appropriate values for silver is in good agreement with the experimental results and are indicated by vertical markers (solid lines) in Fig. 6.5a. The transmission minima are associated with Wood's anomaly.[43] This effect occurs in diffraction gratings when a diffracted order becomes tangent to the plane of the grating. When the order disappears, the light intensity is redistributed among the remaining orders. The intensity in the grazing orders shows a significant increase before being extinguished causing an anomalous reduction of the lower-order intensities.

Of course the anomalous transmission of the hole arrays is present with all metals (Al, Au, Cr, Ag, and similar) but not with non-metals (amorphous Ge for ex.).

[43]Wood RW (1902) Phil Mag. 4:396; Wood RW (1935) Phys Rev. 48:928

Not only is the transmission much higher than expected from classic diffraction theory, it can be greater than the percentage area occupied by the holes, confirming that even the light impinging on the metal between the holes can be transmitted. In other words, the whole periodic structure acts like an antenna in the optical regime.

The enhanced transmission can be explained remembering the properties of photonic crystals. A periodic arrangement of defects on a metal-dielectric interface should exhibit the properties of a one- or two-dimensional "polariton crystal" when the periodicity of the surface structure is comparable with the wavelength of the SPPs propagating on the interface.

The simplest realization of polaritonic crystals is a metallic diffraction grating. Diffraction gratings are widely used for the excitation of surface polaritons on smooth metal films. In this case, surface polaritons are excited on a diffraction grating but propagate on a smooth interface; thus the SPP dispersion relation we have derived remains valid. In the case of a periodically structured film, surface polaritons are both excited by and propagate on a periodic structure. SPP propagation on a periodic structure results in changes in the dispersion relations due to the interaction with periodic surface features. This interaction leads to scattering of SPPs in SPPs as well as scattering of SPPs in light. The increased radiative losses due to the latter process naturally lead to the broadening and spectral shift of the SPP resonance. The former process (SPP-in-SPP scattering) followed by multiple SPP beam interference of the SPPs scattered by periodic features lead to the appearance of SPP band gaps and SPP Bloch waves on a periodically structured surface.[44]

The periodic structure provides the wave-vector conservation (Fig. 6.6a) for surface polariton mode excitation

$$k_{SP} = (\omega/c)\varepsilon_j^{1/2} \sin\theta \mathbf{u}_{xy}\xi \pm p(2\pi/d)\mathbf{u}_x \pm q(2\pi/d)\mathbf{u}_y \tag{6.9}$$

where $\mathbf{u}_{xy}$ is the unit vector in the direction of the in-plane component of the incident light wave vector, $\xi = 1$ for p-polarized (with respect to a surface plane) component of the incident light and 0 for the s-polarized component, $\mathbf{u}_x$ and $\mathbf{u}_y$ are unit lattice vectors of the periodic structure with period d, and p and q are integer numbers corresponding to the different propagation directions of the excited SPPs.

Only the surface polaritons satisfying the Bloch wave condition can be excited and propagate on such a surface. Figure 6.6b shows the band structure of the surface polaritons. For illumination at normal incidence ($\theta = 0$), the only SPPs in the vicinity of even-numbered gaps can be excited directly and correspond to standing SPP Bloch waves on a periodic surface. The periodic structure modifies the local density of electromagnetic states on the metal film interfaces leading to resonant light tunnelling at the wavelengths of the SPP Bloch waves related to one or both film interfaces.

[44] Kitson SC et al (1996) Phys Rev Lett. 77:2670

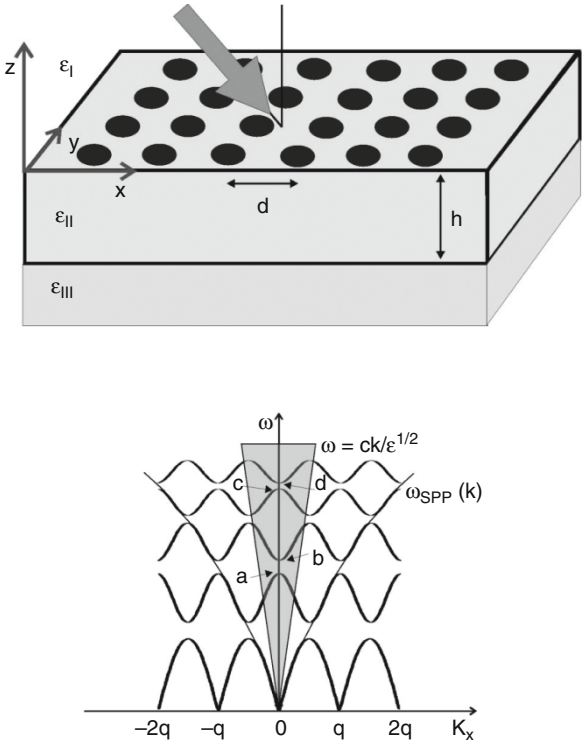


Fig. 6.6 Schematic view of the (**a**) a hole array, (**b**) band structure of surface polaritons on the periodically structured metal surface shown in (**a**). The dispersion of SPPs on a smooth surface $\omega_{SP} = \omega(kS)$ and light $\omega = ck/\varepsilon^{½}$ are shown with thin solid lines. (From S A Darmanyan and A V Zayats[45], Phys. Rev. B67, 035424 (2003) reproduced with permission from AIP)

One and two-dimensional surface polaritonic crystals have been intensively studied.[46]

The extraordinary enhanced optical transmission through a periodic hole array in a metal film can then be explained by resonant light tunnelling involving SPP states formed on one (or both) film interface due to the presence of a periodic structure.[47]

In PBG arrays, the material is passive and transparent at all wavelengths except at the energies within the gap. In the hole arrays, the material plays an active role (trough the plasmons) and it is opaque at all wavelengths except those for which coupling occurs.

Since phase-matching of the incident radiation to SPPs is crucial for transmission enhancement via SPP tunnelling, the same process should occur for a single hole surrounded by a regular array of opaque surface corrugations and this is what

[45]Darmanyan SA, Zayats AV (2003) Phys Rev. B67:035424

[46]Laks B et al (1981) Phys Rev. B23:4965; Glass NE et al (1984) Phys Rev. B29:6548; Barnes WL et al (1996) Phys Rev. B54:6227; Celli V et al (1988) Phys Rev. B37:9089; Watts RA et al (1996) J Mod Opt. 43:1351; Barnes WL et al (1997) J Opt Soc America A14:1654; Tan WC et al (1999) Phys Rev. B59:12661; Kretschmann M et al (2003) Opt Comm. 215:205

[47]Salomon L et al (2001) Phys Rev Lett. 86:1110; Darmanyan S, Zayats AV (2003) Phys Rev. B67:035424

effectively happen.[48] Transmission through an aperture in a regularly patterned surface occurs via tunnelling, leading to an approximately exponential dependence of the transmitted intensity on the thickness t of the metal screen. However, if t is of the order of the skin depth, coupling between SPPs at the front and back interface takes place if the adjacent dielectric media are equal, enabling phase-matching.[49]

In the case of apertures allowing a propagation mode, the transmission can be modulated via resonances of the fundamental mode, controlled by the thickness of the metal film.[50] Periodic surface corrugations around a single slit, significantly increase the transmission and allow $T > 1$ due to excitation of SPPs.[51]

The emission on the exit side of the screen can be controlled via surface patterning obtaining a highly directional emission with narrow beaming angle.[52] The enhancement depends on the array geometry (hole diameter and periodicity), light wavelength, angle of incidence, as well as material of the film.

Applications of extraordinary transmission are in optical filters, photolithography, near field microscopy.[53]

One may have even light transmission through a metal film without any aperture. We have shown in chapter V that interaction between SPPs sustained at the top and bottom interfaces of a metal film can take place, leading to the establishment of coupled bound and leaky SPP modes if the film is thin enough. If both surfaces are modulated via a grating structure, SPPs can be excited on one side of the interface by direct light illumination via grating coupling, tunnel through the film, and be re-emitted on the other side if the period and height of the gratings on both sides are equal. This form of light transmission through a corrugated, unperforated metal film is accompanied by strong energy localization in the grooves of the grating at the input and the exit side.[54]

There is a wealthy of different situations that can be studied.

6.5 Surface Plasmon Circuitry

The quest for reduced size and power consumption of photonic integrated technologies pushes towards the realization of components scaled beyond the diffraction limit of light (smaller than the wavelength).

---

[48]Grupp DE et al (1999) Advanced Materials 11:860; Thio et al (2001) Opt Lett. 26:1972

[49]Degiron A et al (2002) Appl Phys Lett. 81:4327; Shou X et al (2005) Opt Express 13:9834; van der Molen KL et al (2004) Appl Phys Lett 85:4316; Wang et al (2005) Nature 432:376; Barnes WL et al (2004) Phys Rev Lett. 92:107401

[50]Porto JA et al (1999) Phys Rev Lett. 83:2845

[51]Garcia-Vidal FJ et al (2003) Phys Rev Lett. 90:213901

[52]Lezec HJ et al (2002) Science 27:820

[53]Porto et al (2004); Rigneault et al (2005); Luo X, Ishihara T (2004) Appl Phys Lett. 84:4780; Salomon L et al (2001) Phys Rev Lett. 86:1110; Hashizume J, Koyama F (2004) Opt Express 12:6391

[54]Tan WC et al (2000) Phys Rev. B62:11134

The ability of plasmonics to manipulate light well below the diffraction limit has enabled the miniaturization of a range of chip scale optoelectronic devices. Of the many emerging applications, the development of plasmon-enhanced photodetection schemes has witnessed particularly impressive progress. For such devices, the scaling behaviour is very favourable, and some unavoidable, detrimental optical losses in the metals are often compensated by a substantially improved performance in terms of the electrical power consumption, speed of operation, and the possibility to achieve new functionalities with engineered metallic nanostructures. The great promise offered by plasmonics is still in the opportunity in merging photonics and electronics. Electronic circuits provide us with the ability to control the transport and storage of electrons. However, the performance of electronic circuits is rather limited when digital information needs to be sent from one point to another. Photonics offers an effective solution to this problem by implementing optical communication systems based on optical fibres and photonic circuits. Unfortunately, the micrometer-scale bulky components of photonics have limited the integration of these components into electronic chips, which are now measured in nanometers. Plamonics which merge electronics and photonics at the nanoscale, may offer a solution to this size-compatibility problem. In addition the coupling of light from fibres to electronic circuits, in a CMOS compatible technology offers, by means of plasmonics still a great chance application.

Modern nanofabrication and characterization techniques have made possible to structure metal surfaces to steer and control the flow of SPPs and to map the features of that flow. SPPs can be manipulated on a surface in the same way as light is manipulated and directed in three dimensions, making a new two-dimensional optics feasible.[55] Such two-dimensional quasi-optical circuits can be applied, for example, to optical communications and optical computing. SPP based waveguides can transport the same huge bandwidth of information as in conventional photonics and yet not be limited by diffraction to submicron cross sections. Surface plasmons can also serve as a basis for design, fabrication, and characterization of sub-wavelength waveguide components[56] as filters and sensors, waveguides and complex circuits that will be able to carry optical signals and electric currents.[57]

[55]Han Z, Bozhevolny SI (2013) Rep Prog Phys. 76:016402

[56]Quinten M et al (1998) Opt Lett. 23:1331; Charbonneau R et al (2000) Opt Lett. 25:844; Nikolajsen T et al (2003) Appl Phys Lett. 82:668; Krenn JR et al (2002) Europhys Lett. 60:663; Maier SA et al (2003) Nat Mat. 2:229; Murray WA et al (2004) Phys Rev. B69:165407; Maier SA et al (2004) Appl Phys Lett. 84:3990; Maier SA et al (2005) Appl Phys Lett. 86:071103; Berini P et al (2005) J Appl Phys. 98:043109; Novikov IV, Maradudin A (2002) Phys Rev. B66:035403; Pile DFP, Gramotnev DK (2004) Opt Lett. 29:1069; Pile DFP, Gramotnev DK (2004) Appl Phys Lett. 85:6323; Bozhevolnyi SI et al (2005) Phys Rev Lett. 95:046802; Bozhevolnyi SI et al (2006) Nature 440:508

[57]Barnes WL et al (2003) Nature 424:824; Nomura W et al (2005) Appl Phys Lett. 86:181108; Bozhevolnyi SI (ed) (2009) Plasmonic nanoguides and circuits. Pan Stanford Publishing

Plasmon circuits first convert light into SPPs, which then propagates and are processed by logic elements before being converted back into light. To build such a circuit a variety of components are needed: waveguides, switches, couplers, etc.

The use of SPPs helps us to concentrate light in sub-wavelength structures. Concentrating light in this way leads to an electric field enhancement that can be used to manipulate light-matter interactions and boost non-linear phenomena.

We have seen that the real part of the wave-vector k of the plasmon increases as the plasma frequency is approached and the extension of the evanescent field becomes more and more confined to the metal. The imaginary part of k determines the distance over which the plasmon can travel along the surface, that is the propagation length L. The fact that L decreases as the plasmon becomes more confined is a fundamental feature inherent in all plasmon modes. Stronger plasmon confinement pushes the field closer to the metal, which results in stronger absorption and therefore shorter propagation. To construct a robust plasmon circuit therefore, a compromise between confinement and propagation must be found. At telecommunication wavelengths, around 1.5 μm, L can reach values close to 1 mm at the interface of metals such as silver. Such length is more than sufficient to make a circuit that contains several devices.

A large number of simulations of many different components exist and also the possibility to guide plasmon polaritons at 2D metal corners has been studied.[58]

A thin metal stripe embedded inside a dielectric increases the propagation length[59] in the so-called long-range SPPs. With these configurations a number of elements have been constructed with also more complex elements like directional couplers.[60]

As an example, Charbonneau et al constructed, some basic elements using gold films of several widths (2, 4, 6, and 8 μm), 25 mm thick and 7 mm long, deposited on SiO_2 and covered with an index-matched polymer. All structures operated at 1550 nm and were butt-coupled with an optical fibre. The output of the elements was examined with a microscope objective.

The output of the elements was examined with a microscope objective.

SPP-based components such as mirrors, beam-splitters, interferometers, that are significantly smaller than the propagation length have been constructed.[61]

It is possible to have plasmons travelling at the bottom of V shaped grooves (the so-called Channel Plasmon Polaritons CPPs) and this geometry appears promising. Although CPPs propagate over shorter distances than LRSP (long range surface plasmons), they provide a better confinement which allows for structure designs with sharper turn, as for example, curvatures.[62]

[58]Yan M, Qiu M (2007) J Opt Soc America B24:2333

[59]Boltasseva A et al (2005) J Lightwave Techn. 23:413

[60]Boltasseva A, Bozhevolnyi I (2006) IEEE J Sel Top QE QE-12:1233

[61]Ditlbacher H et al (2002) Appl Phys Lett. 81:1762

[62]Bozhevolnyi SI et al (2006) Nature 440:508; Volkov VS et al (2007) Nano Lett. 7:880

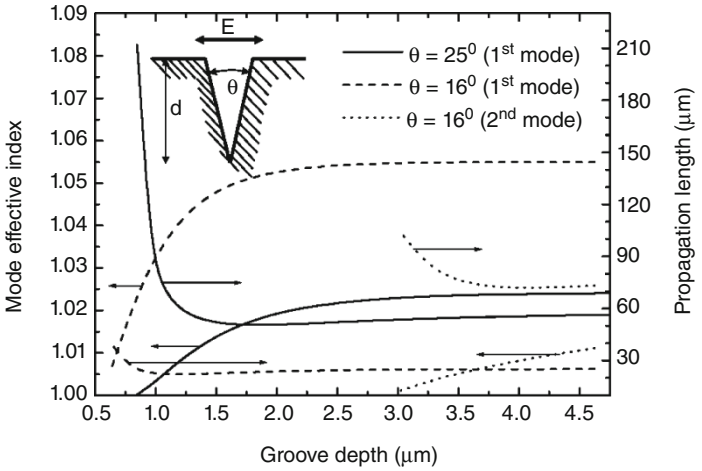


Fig. 6.7 The effective indices of CPP modes and their propagation lengths as a function of groove depth d for different groove angles θ. The insert shows the groove configuration and dominant orientation of CPP electric field. (From S I Bozhevolnyi et al, Phys. Rev. Lett. 95, 046802 (2005) reproduced with permission from AIP

Figure 6.7 shows the simulation[63] of the CPP in a V-shaped gold guiding structure at 1.55 μm. The CPP effective index increases while the propagation length decreases with decrease of the grove angle θ. The CPP mode index determines the mode confinement, since the penetration depth in air (above the surface) is given by

$$d_{pen} = (\lambda/2\pi)(n_{eff}^2 - 1)^{1/2} \tag{6.10}$$

where

$$n_{eff} = (\lambda/2\pi)\mathrm{Re}(\beta) \tag{6.11}$$

so that a larger effective index n_{eff} corresponds to a smaller penetration depth and thereby a better field confinement in depth. In width, the CPP mode is confined within the grove walls. Furthermore, the CPP propagation length starts to rapidly increase when the groove depth decreasing approaches a certain cut-off value. Since, at the same time, the CPP effective index approaches that of air, the increase of propagation length signifies the tendency of the CPP mode field being extended progressively outside the groove. This behaviour has been confirmed experimentally.[64]

[63] Bozhevolnyi SI et al (2005) Phys Rev Lett. 95:046802

[64] Bozhevolnyi SI et al (2006) Nature 440:508

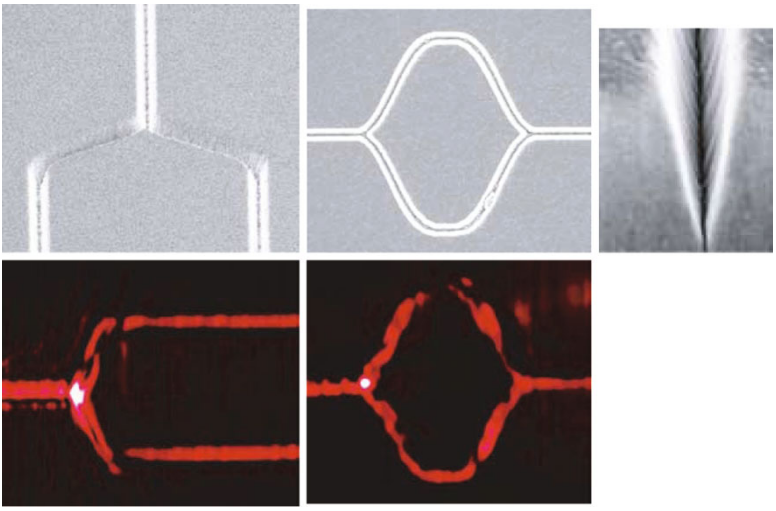


Fig. 6.8 Upper left SEM image of the Y-splitter and on the right the Mach-Zehnder. At the extreme right the groove profile. Lower part near field optical (1600nm) SNOM images of the Y splitter and the Mach-Zehnder interferometer. (From S I Bozhevolnyi et al., Nature 440, 508 (2006) (c) Nature)

Bozhevolnyi et al.,[65] designed various sub-wavelength waveguide structures which included V-grooves with angles close to 25° and depths of 1.1–1.3 μm fabricated by focused ion-beam milling in a 1.8 μm thick gold layer deposited on a substrate of fused silica, Y splitters and Mach-Zehnder interferometers (Fig. 6.8) and waveguide ring resonator.

Channel plasmon polaritons trapped at the bottom of grooves engraved in a thick metal film which strongly confines the plasmon field, belong to a family of waveguides whose modes are supported by a narrow gap between two metal surfaces, in contrast to the modes of thin metal films. In gap waveguides that confine the SPPs on the gap-width between two parallel metallic plates,[66] the plasmon modes on opposite walls of the gap overlap, couple, and form two new modes with opposite symmetry. However, the only mode that survives for all gap widths, the so called gap plasmon or GSP is the one whose normal (transverse) electric field component maintains its sign across the gap.

Using CPPs a number of devices were built. A nano-focusing system was described by Volkov et al.[67] A simple add-drop multiplexer based on a waveguide-ring resonator WRR ADM and a compact Bragg grating.[68] The multiplexer is shown in Fig. 6.9.

[65]Bozhevolnyi SI et al (2005) Phys Rev Lett. 95:046802; Bozhevolnyi SI et al (2006) Nature 330:508

[66]Tanaka K, Tanaka M (2003) Appl Phys Lett. 82 :1158

[67]Volkov VS et al (2009) Nano Lett. 9:1278

[68]Volkov VS et al (2007) Nano Lett. 7:880

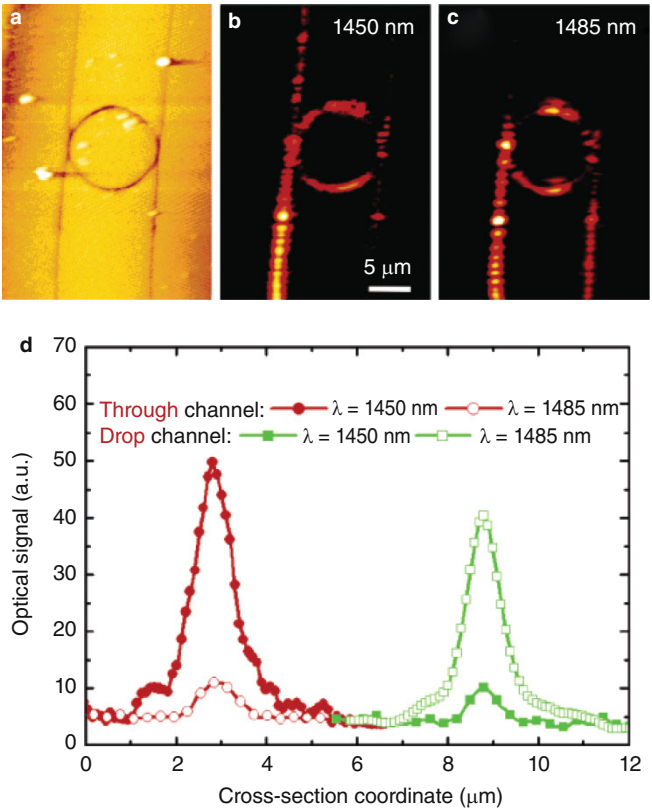


Fig. 6.9 Pseudocolor (**a**) topographical and (**b** and **c**) near field optical images of the WRR ADM at (**b**) 1450 nm and (c) 1485 nm. (**d**) Normalized cross-sections obtained from the optical images shown in (**b**) and (**c**) for two different output (through and drop) channels with the signal being normalized to be at the same level in the input channel. (From Volkov et al. Nano Letters, 7, 880 (2007) (c) NanoLetters

Pyayt et al.[69] have made multiple silver nanowire SPP waveguide with polymer optical waveguides for nanoscale confinement and guiding of light on a chip.

A number of optical elements have been made by Ditibacher et al.[70] Figure 6.10 shows a mirror obtained using an ensemble of nanoparticles arranged to parallel lines to act as a Bragg reflector.

Several review papers describe the status of the art for realizations and theory.[71]

[69]Pyayt AL et al (2008) Nat Nanotechn. 3:660

[70]Ditibacher H et al (2003) Appl Phys Lett. 81:1762

[71]Zhang J, Zhang L, Xu W J. Phys. D; Zhang J, Zhang L, Xu W (2012) Appl Phys. 45:113001; Barnes WL, Dereux A, Ebbesen TW (2003) Nature 424:824; Ebbesen TW, Genet C, Bozhevolnyi SI (Mar 2008) Physics Today. p. 44; Zayats AV, Smolyaninov I (2003) J Opt A Pure Appl Opt. 5:S16; Sorger VJ et al (2012) MRS Bulletin 37:728

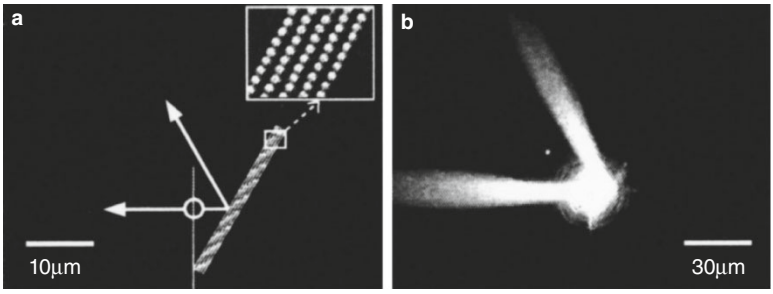


Fig. 6.10 (**a**) SEM image of the SPP Bragg reflector built up from individual nanoparticles (diameter 140 nm, height 70 nm); the inset shows a 7× magnified image of the area marked by the white box. The circle defines the focus position and the arrows indicate the direction of SPP propagation (**b**) corresponding fluorescence image. (From Ditibacher et al., Appl. Phys. Lett. 81, 1762 (2003) reproduced with permission from AIP)

6.6 Plasmon Lasers and SPASER

The smaller is a laser, the smaller is the required operating power, and potentially it can be switched on and off faster or modulated at higher frequencies than ordinary macroscopic devices.[72] With the increased miniaturization of all electronic devices, lasers with the dimension of wavelength had become too large for some applications.

Fortunately, with the advances in technology, in the 1990s, construction of objects of dimensions of a few nanometers has become possible. Accordingly, what previously appeared impossible: the construction of a resonant cavity of the dimension of the visible wavelength became a routine affair and microscopic lasers – sometimes called nanolasers – that can reach the diffraction limit (that is have dimensions of $\lambda/2$) have been built.

Several micro- and nano-lasers have been demonstrated based on dielectric cavities, such as photonic crystals,[73] microdisks,[74] micropillars,[75] and nanowires.[76]

[72]Altug H, Englund D, Vuckovic J (2006) Nature Phys. 2: 484

[73]Painter G et al (1999) Science 284:1819; Park HG et al (2004) Science 305:1444; Nozaki K et al (2007) Opt Exp. 15:7506; Noda S (2010) J Opt Soc America B27:B1

[74]Tamboli AC et al (2007) Nature Phot. 1:61; Liu L et al (2010) Nature Photon. 4:182; Song Q et al (2009) Appl Phys Lett. 94:061109; Ide T et al (2005) Opt Exp. 13:1615; Zang ZY et al (2007) Appl Phys Lett. 90:111119

[75]Ulrich SM et al (2007) Phys Rev Lett. 98:043906; Reitzenstein S et al (2008) Appl Phys Lett. 93:061104; Albert F et al (2010) Appl Phys Lett. 97:101108

[76]Huang MH et al (2001) Science 292:5523; Johnson JC et al (2002) Nature Mat. 1:106; Duan X et al (2003) Nature 421:241; Chin AH et al (2006) App Phys Lett. 88:163115; Seo MK et al (2008) Nano Lett. 8:4534; Xia F et al (2008) Nature Nanotech. 3:609; Zimmler MA et al (2008) Appl Phys Lett. 93:051101

All these dielectric lasers have however volume mode or physical dimensions too large to cope with electronic integrated circuits.

In theory it is possible to have a better confinement of light with the use of metals to form the laser resonator. The metal may form either strong compact mirrors, able to confine light to about the size of the diffraction limit, or the light may interact strongly with the free electrons in the metal, being guided in the form of surface plasmon polaritons (SPPs) at the interface between a metal and a dielectric material such as air. This way plasmon nanolasers smaller than the 3-D diffraction limit can be built.

The minimum size scales of structures employing this form of light confinement is related to the penetration depth of light into the metal, which is typically of the order of tens of nanometers. Initially many researchers doubted if it would be possible to overcome the losses in plasmonic or metallic waveguides and cavities with the currently available optical gain materials. The metal's conduction electrons, which oscillate in synchronism with the optical field, dissipate energy through collisions with the metal's atomic lattice: a process that leads to high optical losses, so surface plasmons can travel only short distances, and metal mirrors have higher losses than dielectric ones. However, the possibility of using gain in metallic waveguides was considered[77] and microscopic lasers that can reach the diffraction limit, based on metal-clad cavities have been successfully reported notwithstanding the high losses.

The first report of lasing in metallic nano-structures occurred in 2007[78] with what was dubbed the "gold-finger" laser. It was the first experimental attempt using metals to confine the optical energy to lasing using a round semiconductor pillar encapsulated in a dielectric and then gold, pumped electrically. Successively the structure was repeated as a cylindrical semiconductor pillar with a height of 1.6 μm and a diameter of about 250 nm, consisting of an InGaAs active medium sandwiched by InP pillar sections, surrounded by a thin insulating Si_3N_4 layer and followed by a silver cladding. The metallic cavity acted as the n-contact itself, and a p-contact was laterally inserted (Fig. 6.11).[79]

The pillar had an InGaAs active region height of 300 nm. The InGaAs heterostructure in the pillar and the metal formed a resonator with an optical mode trapped on the InGaAs gain medium (Fig. 6.11b) in a region with dimensions approximately one half the wavelength of light in the medium, though it was diffraction limited because of its nonplasmonic nature. The large resistive losses associated with the metal required cryogenic temperatures to obtain sufficient gain in the InGaAs to overcome losses in the resonator and achieve lasing at a wavelength of approximately 1400 nm. This small laser was electrically pumped with a

[77] Nezhad MP et al (2004) Opt Express 12:4072

[78] Hill MT et al (2007) Nature Photon. 1:589

[79] Hill TM (2010) J Opt Soc Am. B27:B36

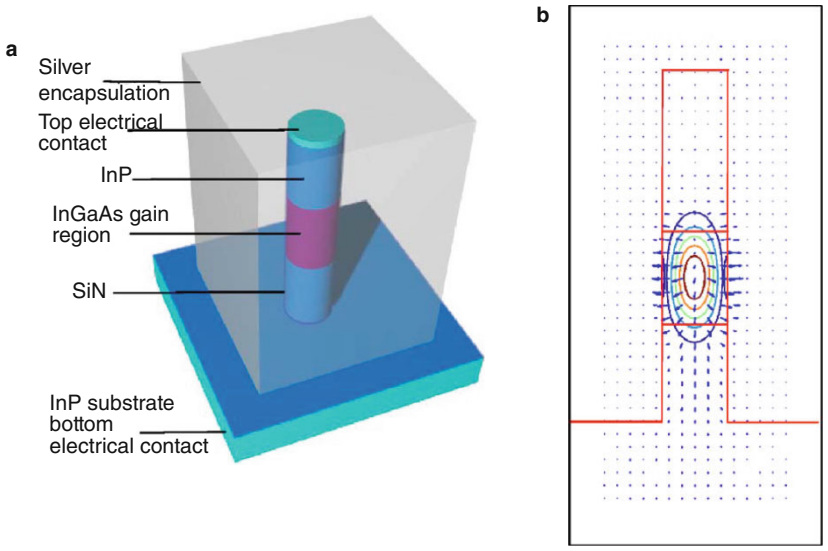


Fig. 6.11 (**a**) A semiconductor heterostructure pillar is encapsulated in an insulator, then a low-optical-loss metal. Light escapes through the bottom of the pillar and the substrate. (**b**) The optical mode trapped in the InGaAs as shown by the electric field squared contour lines. (From T M Hill, J. Opt. Soc. America B27, B36 (2010) © Optical Society America)

threshold current of 6 μA at 77K. Later, a nanolaser showing plasmonic character with one-dimensional confinement was demonstrated,[80] a distributed feedback laser[81] and a metallic-cavity nanolaser operating at room temperature under cw electric injection was built.[82] The structure of this laser is shown in Fig. 6.12.

It consist of an InP/InGaAs/InP pillar with a 20 nm Si_3N_4 layer on all four sides. The device layer forms a three-layer sandwich waveguide in the vertical (z) direction, confining light mostly in the GaAs. This pillar is then encapsulated in silver from all four sides and the top, forming a metal-insulator-semiconductor-insulator-metal waveguide in the horizontal (x and y) directions. Emission is collected from the backside of the substrate by an objective lens.

The light output versus current curve is shown in Fig. 6.13 The full width at half maximum of the lasing mode shows a rapid decrease with increasing current, typical of laser action.

Several devices were built of different dimensions. Figure 6.13 refers to a device with a volume of $0.42\lambda^3$.

[80] Hill MT et al (2009) Opt Express 17:11107

[81] Marell MJH et al (2011) Opt Express 19:15109

[82] Ding K et al (2012) Phys Rev. B 85:041301–1–5

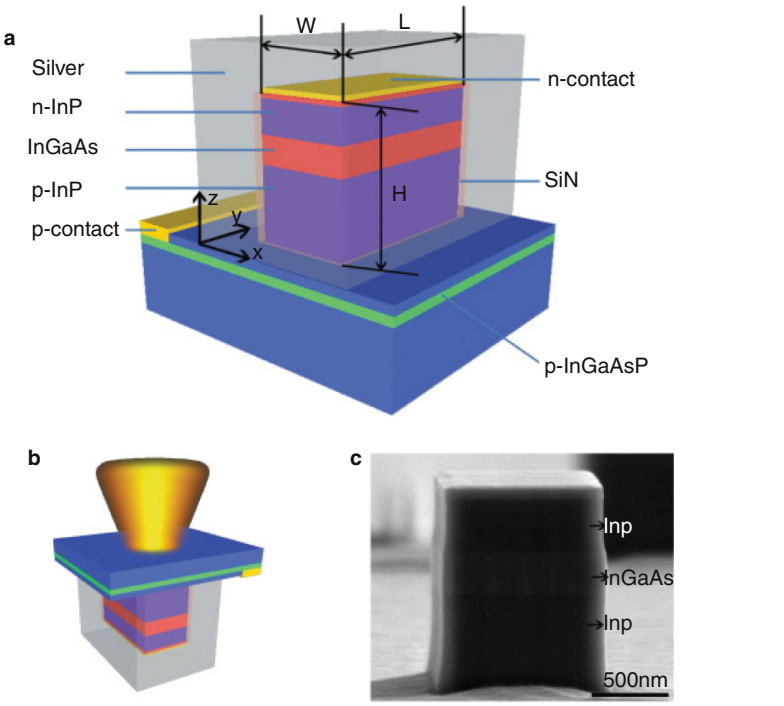


Fig. 6.12 (**a**) Semiconductor pillar of a rectangular cross-section encapsulated in silver to form a metallic cavity. (**b**) Schematic of the laser structure inverted showing laser light emitted from the backside of the substrate. (**c**) Scanning electron microscope image showing the semiconductor pillar before SiN and silver coating. (From K Ding et al., Phys. Rev. B85, 041301-1-5 (2012) reproduced with permission AIP)

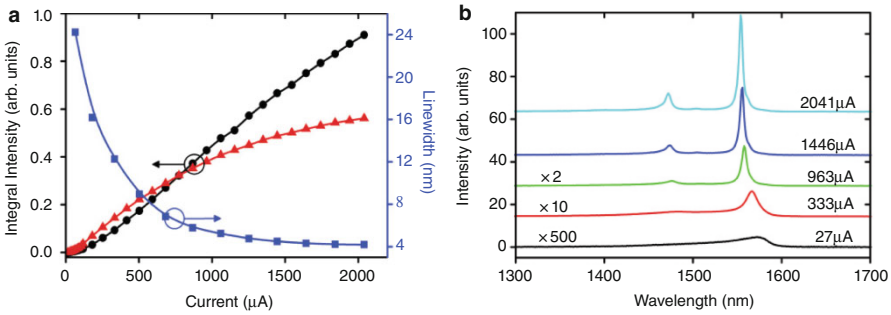


Fig. 6.13 (**a**) Linewidth (squares), integrated lasing mode intensity (circles) and integrated spontaneous emission intensity (triangles) as a function of dc current at 293K. The threshold is estimated about 1000 μA (**b**) spectra at different dc current. (From K Ding et al., Phys. Rev. B85, 041301-1-5 (2012) reproduced with permission AIP)

A number of other nanolasers were made from metal semiconductor structures lasing in the near infrared. A metal-dielectric encapsulated semiconductor heterostructure showed room temperature operation for an optically pumped device that was less than a free space wavelength of light in all three dimensions.[83]

Plasmon modes have been studied in several geometries as micro- or nano-disk cavities having one side coated with metal[84] or a small patch of metal-semiconductor-metal sandwich at cryogenic temperatures[85] or a plasmon laser with total internal reflection cavity.[86]

Oulton et al[87] showed lasing in a section of a hybrid dielectric-plasmonic waveguide to overcome the limitations due to losses. The waveguide was constructed from a flat metal substrate (silver) coated with a thin dielectric layer on which a thin high-index semiconductor cadmium sulfide nanowire was placed. A hybrid plasmon mode was concentrated in the insulator gap of 5–10 nm with its tail overlapping with the semiconductor gain. In this approach, the electromagnetic field is lifted from the metal into the dielectric gap, resulting in low loss operation, yet maintaining the plasmonic nature of high confinement. The optical mode propagating in the silver-dielectric-semiconductor waveguide has a significant amount of its energy squeezed into the thin dielectric gap between the nanowire and the metal substrate, leading to a highly localised mode which has an area significantly below the diffraction limit. The nanowire forms a Fabry-Perot cavity with the plasmonic modes resonating between its two ends which are a few micrometers apart. Using short optical pulses to pump the semiconductor nanowire, lasing was shown in the Fabry-Perot cavity at low temperature in the blue part of the visible spectrum.

6.6.1 Spaser

A way to obtain shorter confinement is to make use of surface plasmons. Suitably coupling light to plasmons, one may built a true *nanolaser*, that is a miniature laser of nanoscale dimensions, because a plasmon has a wavelength much shorter than light in vacuum. To overcome the ohmic losses that at optical frequencies make a challenge to realize lasers making use of surface plasmons, a source of energy must be provided.

[83]Nezhad MP et al (2010) Nature Photon. 4:395

[84]Perahia R et al (2009) Appl Phys Lett. 95:201114; Kwon SH, Kang JH, Kim SK, Park HG (2011) IEEE J Quantum Elect. 47:1346–1353

[85]Yu K, Lakhani A, Wu MC (2010) Opt Express 18:8790

[86]Oulton PF et al (2008) Nat Photonics 2:496; Sorger VJ et al (2011) Nature Commun. 2:331; Ma RM et al (2011) Nat Mater. 10:110

[87]Oulton RF et al (2009) Nature 461:629

Bergman and Stockman in 2003[88] described a system they dubbed *spaser*. *Spaser* is the acronym for Surface Plasmon Amplification by Stimulated Emission of Radiation, and is a nanoplasmonic counterpart of the laser in which photons are replaced by surface plasmons. A spaser consists of a metal nanoparticle, which plays the role of the laser cavity (resonator), and the gain medium which is an active medium with an inverted population and supplies energy to the lasing.

The physical principle is the amplification of surface plasmons at the expenses of the excited material, typically a semiconductor or a dye.

An inherent limitation of a traditional laser is that the size of the laser cavity in the propagation direction is at least half wavelength and practically more than that even for the smallest developed lasers. In a true spaser, this limitation is overcome. The spasing modes are surface plasmons (SPs) whose localization length is on the nanoscale.[89]

One of the simplest and potentially most promising types of nanoparticles to function as a spaser resonator is a metal–dielectric nanoshell. Such nanoshells have been introduced by Naomi Halas et al. and have since found a very wide range of applications.[90] To better understand its working we follow the Stockman papers and show in Fig. 6.14 the field around a metallic nanoparticle (Fig. 6.14a, b) together with two possible spaser realisations (Fig. 6.14c,d). In Fig. 6.14c the gain medium is surrounding the metallic nanosphere and the plasmonic mode develops at the interface between the nanosphere and the gain medium. In Fig. 6.14d the gain medium is inside the metallic nanosphere. Let us consider the geometry in which a metallic (gold) sphere is covered with a suitable material such as a semiconductor or a dye. Upon excitation of the dye with a suitable pumping laser a transfer of an optical excitation to a plasmon should occur. A plasmon mode in the shell is therefore feed at the expenses of the optically excited dye.

Figure 6.15 shows the physical operation. The gain medium, for example a semiconductor, is excited and then decays on an intermediate level which has an energy very near to the plasmon energy.

The gain medium chromophores may be semiconductor nanocrystal quantum dots,[91] dye molecules,[92] rare-earth ions,[93] or electron-hole excitations of an unstructured semiconductor as used by Hill.[94]

---

[88]Bergman DJ, Stockman MI (2003) Phys Rev Lett. 90:027402, see also Stockman MI (2011) Phys Today 64:39; Stockman MI, Bergman DJ (2009) Surface plasmon amplification by stimulated emission of radiation (spaser), US Patent 7, 569:188; Li K, Li X, Stockman MI, Bergman DJ (2005) Phys Rev. B 71:115409; Bergman DJ, Stockman MI (2004) Laser Phys. 14:409; Stockman MI (2011) Opt Express 19:22029

[89]Stockman MI, Faleev SV, Bergman DJ (2001) Phys Rev Lett. 87:167401

[90]Averitt RD, Sarkar D, Halas NJ (1997) Phys Rev Lett. 78:4217

[91]Plum E, Fedotov VA, Kuo P, Tsai DP, Zheludev NI (2009) Opt Express 17:8548

[92]Seidel J, Grafstroem S, Eng L (2005) Phys Rev Lett. 94:177401; Noginov MA et al (2008) Phys Rev Lett. 101:226806

[93]Gordon JA, Ziolkowski RW (2007) Opt Express 15:2622

[94]Hill MR et al (2007) Nat Phot. 1:589; Hill MT et al (2009) Opt Express 17:11107

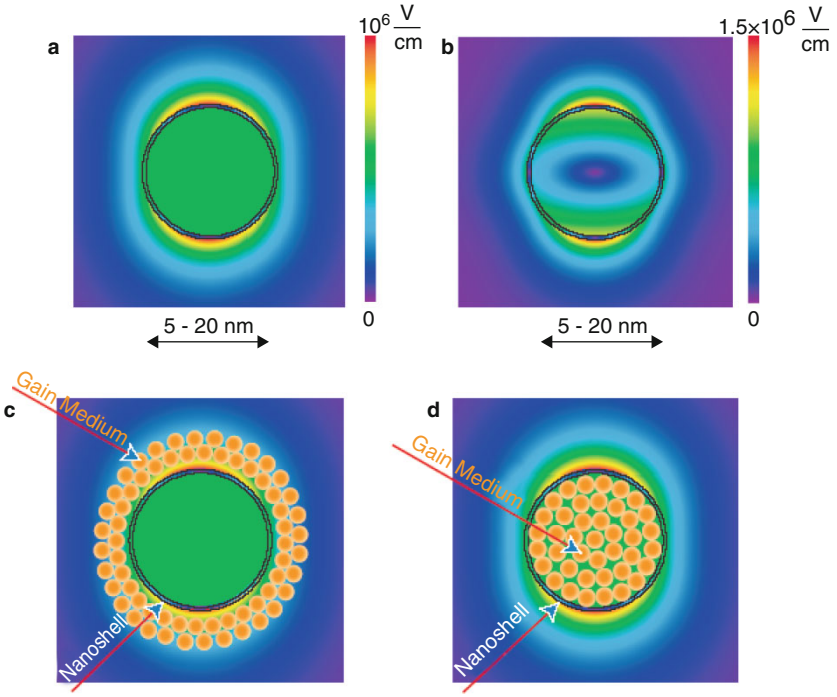


Fig. 6.14 Schematic of spaser geometry, local fields, and fundamental processes leading to spasing. (**a**) Nanoshell geometry and the local optical field distribution for one SP in an axially-symmetric dipole mode. The nanoshell has aspect ratio h = 0:95. The local field magnitude is color-coded by the scale bar in the right-hand side of the panel. (**b**) The same as (**a**) but for a quadrupole mode. (**c**) Schematic of a nanoshell spaser where the gain medium is outside of the shell, on the background of the dipole mode field. (**d**) The same as (**c**) but for the gain medium inside the shell. (From M Stockman, Opt. Express 19, 22029 (2011) © Optical Society America

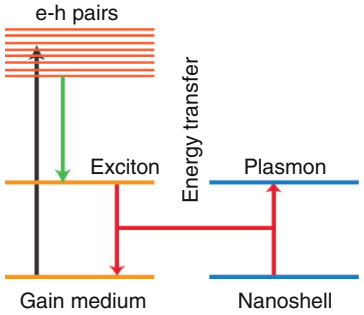


Fig. 6.15 Schematic of the spasing process. The gain medium is excited and population-inverted by an external source, as depicted by the black arrow, which produces electron-hole pairs in it. These pairs relax, as shown by the green arrow, to form the excitons. The excitons undergo decay to the ground state emitting SPs into the nanoshell. The plasmonic oscillations of the nanoshell stimulates this emission, supplying the feedback for the spaser action. (From M Stockman, Nature Photonics 2, 327 (2008) © Nature Photonics

Let us consider quantum dots. The pump excites electron-hole pairs in the chromophores (Fig. 6.15), as indicated by the vertical black arrow, which relax to form excitons. The excitons constitute the two-level systems that are the donors of energy for the SP emission into the spasing mode. In vacuum, the excitons would recombine emitting photons. However, in the spaser geometry, the photoemission is strongly quenched due to the resonant energy transfer to the surface plasmon modes, as indicated by the red arrows in the panel.

The plasmons already in the spaser mode create the high local fields that stimulate more emission to this mode by the gain medium, which is the feedback mechanism. If this feedback is strong enough, and the life time of the spaser SP mode is long enough, then an instability develops leading to the avalanche of the SP emission in the spasing mode and spontaneous symmetry breaking, establishing the phase coherence of the spasing state.

The theory of the device, first presented by Bergman and Stockman, has been further developed.[95]

When the device works, a narrowing of the emission line-width occurs and a collimated beam may be produced.

As Stockman pointed out, one of the advantages of the spaser compared with existing sources of local fields, which also sets it apart from the laser, is that it can generate dark modes that do not couple to the far-zone optical fields. In other words, a spaser generates coherent, strong local fields, but does not necessarily emit photons. This is a potentially great technological advantage because it offers a source of nano-localized optical fields that does not emit any background radiation. The source can still act on molecules in its near field and excite their radiation (such as fluorescence and Raman effects) as conventional nano-optical sources do.

Spaser effect was observed experimentally[96] in 2009, using 44-nm-diameter nanoparticles with a gold core and a dye-doped silica shell containing the organic dye Oregon Green 488 (OG-488), providing for gain (Fig. 6.16).

Emission was expected at 525 nm with a quality factor Q = 14.8. The measured extinction spectrum, shown in Fig. 6.17, is dominated by the surface plasmon resonance band at ~520 nm. The samples were pumped at 488 nm with 5 ns pulses. Once the pumping energy exceeded a critical threshold value, a narrow peak appeared at 531 nm.

The emission wavelength corresponds to the lower order localised plasmon mode of the gold sphere.

Although this nanoparticle approach provides the ultimate scaling down in all three dimensions, its optical mode extends appreciably outside the structure, and electrical connections are difficult to implement.

[95]Stockman MI (2008) Nature Photonics 2:327; Stockman MI (2010) J Optics 12:024004; Stockman MI (2011) Phys Rev Lett. 106:156802; Stockman MI (2011) Phil Trans Roy Soc. A369:3510; Stockman MI (2011) Opt Express 19:22029; Andrianov ES et al (2011) Opt Lett. 36:4302; Stockman MI (2011) Opt Express 19:22029

[96]Noginov MA et al (2009) Nature 460:1110

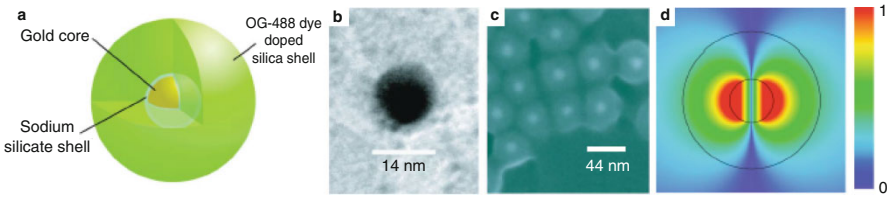


Fig. 6.16 Spaser design. **a** Diagram of the hybrid nanoparticle architecture (not to scale), indicating dye molecules throughout the silica shell. **b** Transmission electron microscope image of Au core. **c** Scanning electron microscope image of Au/silica/dye core–shell nanoparticles. In **d** spacer mode (in false colour), with $\lambda = 525$ nm and $Q = 14.8$; the inner and the outer circles represent the 14-nm core and the 44-nm shell, respectively. The field strength colour scheme is shown on the right. (From M A Noginov et al., Nature 460, 1110 (2009) © Nature)

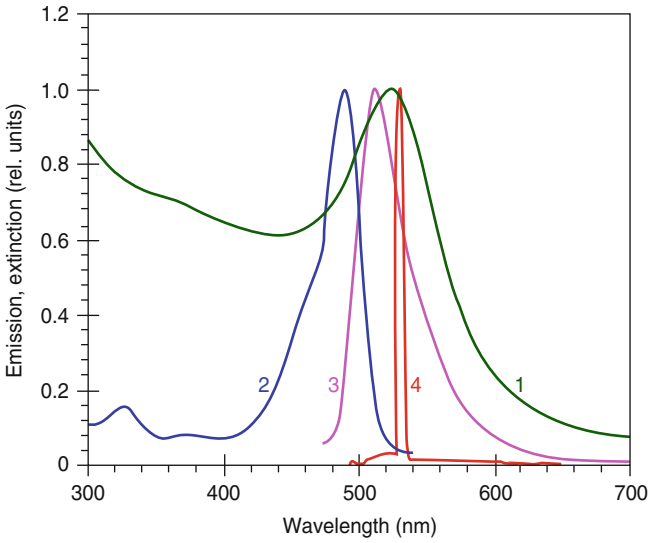


Fig. 6.17 Spectroscopic results. Normalized extinction (1), excitation (2), spontaneous emission (3), and stimulated emission (4) spectra of Au/silica/dye nanoparticles showing the peak of lasing emission at 530 nm. (From M A Noginov et al., Nature 460, 1110 (2009) © Nature)

A number of surface plasmon polariton (SPP) spasers with different geometries have been experimentally observed.[97] A beautiful planar spaser[98] is shown in Fig. 6.18. The field distribution in the structure, as shown in Fig. 6.18b makes

[97]Sorger VJ, Zhang X (2011) Science 333:709; Ma RM et al (2011) Rev. Sci. Instrum. 82:033107; Ma M et al (2013) Laser Photonics Rev. 7:1; Suh JY et al (2012) Nano Lett. 12:5769; Liu Y et al (2011) Opt Lett. 36:1296; Flynn RA et al (2011) Opt Express 19:8954; Wu CY, Kuo CT, Wang CY, He CL, Lin MH, Ahn H, Gwo S (2011) Nano Lett. 11:4256

[98]van Beijnum F et al (2013) Phys Rev Lett. 110:206802

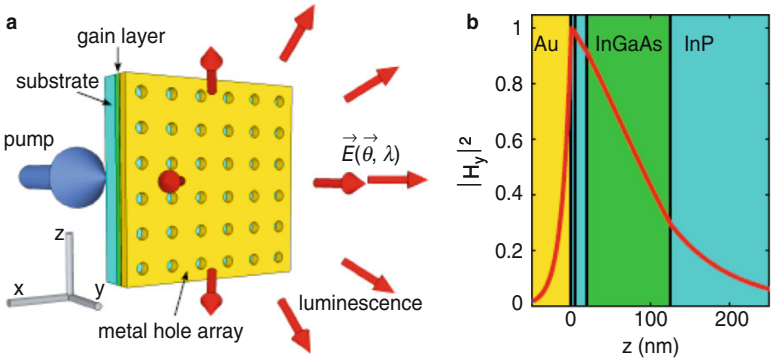


Fig. 6.18 (**a**) A semiconductor layer in close proximity of a metal hole array is pumped optically; (**b**) The simulation of the absolute value squared of the magnetic field of the surface plasmon mode. (From F van Beijnum et al., Phys. Rev. Lett. 110, 206802 (2013) reproduced with permission AIP)

clear how the evanescent tail of the Plasmon in the semiconductors extend so much as to gain energy from them. Figure 6.18a shows the output spectrum as a function of pump power. At 40 mW pump power the peak at 1480 nm increases dramatically and the presence of a threshold is clearly shown in the plot of Fig. 6.19b.

A spaser emitting continuously in the green has been made by Yu-Jung Lu et al.[99] The laser plasmonic nanocavity is formed between a smooth silver film and a single optically pumped nanorod consisting of epitaxial gallium nitride shell and an indium gallium nitride core (gain medium). The atomically smooth silver layer decreases the modal volume, which reduces surface plasmonic losses.

An electrical injection sub-wavelength metallic-cavity semiconductor laser operating cw at room temperature was made by Ding et al.[100]

Other proposals to realise spasers have been presented.[101] Stockman has suggested graphene as a suitable material and has discussed an electric spaser in the extreme quantum limit.[102]

Review papers on metallic and plasmonic nanolasers,[103] photonic crystal lasers,[104] plasmon lasers[105] have been published.

[99]Lu YJ et al (2012) Science 337:450

[100]Ding K et al (2013) Opt Express 21:4728

[101]Zheludev NI, Prosvirnin SL, Papasimakis N, Fedotov VA (2008) Nature Photon. 2:351

[102]Apalkov V, Stockman MI (2013) Graphene spaser, arXiv:1303.0220v1; Li D, Stockman MI (2013) Phys Rev Lett. 110:106803

[103]Hill MT (2010) J Opt Soc Am. B27:B36

[104]Noda S (2010) J Opt Soc Amer. B27:B1

[105]Ma RM et al (2013) Laser Photonics Rev. 7(1):1

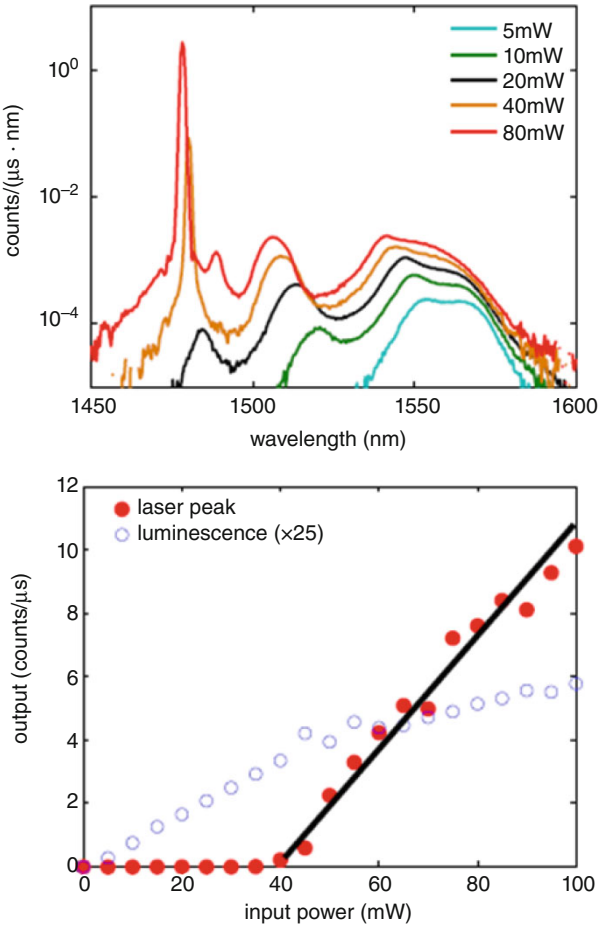


Fig. 6.19 (**a**) Upper diagram, luminescence spectra as a function of pump power. Increasing pump power the bandwidth of the luminescence increases until the device starts lasing. (**b**) Lower diagram shows that the power in the lasing peak exhibits a clear threshold at 40 mW. (From F van Beijnum et al., Phys. Rev. Lett. 110, 206802 (2013) reproduced with permission AIP)

6.7 Plasmons for Solar Cells

Solar cells are solid-state electrical devices that converts the energy of light directly into electricity by the photovoltaic effect. Most solar cells are based on crystalline silicon wafers with thicknesses between 180 and 300 µm, and most of their price is due to the costs of silicon material and processing. Because of this, there is a great interest in thin-film solar cells, with film thicknesses in the range 1–2 µm.

A large thickness is needed for increasing the absorption of light; on the other hand high-efficiency solar cells must have minority carrier diffusion lengths several times the material thickness for all photocarriers to be collected, a requirement that is most easily met for thin cells.

Conventional light trapping schemes use wavelength scale surface texturing on the front or back of the cell to enhance light absorption. These structures are too large to be used with extremely thin films. Metal nanostructured thin films have the potential to confine and guide incident sunlight into sub-wavelength thickness absorber layer volume due to excitation of SPPs, and the photovoltaic conversion efficiency of solar cells can be enhanced by incorporation of small metal nanostructures.

Both localized surface plasmons excited in metal nanoparticles and surface plasmon polaritons propagating at the metal/semiconductor interface can be used. A number of proposals and simulations have shown that plasmonic structures can offer a number of possibilities for reducing the physical thickness of photovoltaic absorber layers keeping their optical thickness constant. Metallic nanoparticles can be used as sub-wavelength scattering elements to couple and trap freely propagating plane waves from the Sun into an absorbing semiconductor this film, by folding the light into thin absorber layer.

Light scattering from a small metal particle embedded in a homogeneous medium is nearly symmetric in the forward and reverse direction. However when the particle is placed close to the interface between two dielectrics, light will be scattered preferentially into the dielectric with the larger permittivity. The scattered light may fill the entire film region, can be back-scattered by the back surface and remain trapped into film. The shape and size of metal nanoparticles are key factors determining the in-coupling efficiency.[106]

Plasmon enhanced light absorption has been demonstrated to increase the cell photocurrent in many types of cells made with crystalline[107] or poly[108] Si thin-film solar cells. By properly engineering of these structures, light can be concentrated and coupled into a thin semiconductor layer to increase light absorption. After many theoretical discussions, first experimental results have been obtained by a number of researchers.

Park et al.[109] have applied silver nanoparticles on the rear side of polycrystalline silicon thin film solar cells and investigated the optimum surface conditions for maximizing plasmonic enhanced light absorption in the cells.

Because the environment affects the surface plasmon resonance wavelength, the scattering angle, the scattering cross-section, and the coupling efficiency, the nanoparticles were embedded in three different environments, thermal silicon dioxide (SiO_2), native SiO_2, and oxide-free silicon surface. In all cases an enhancement of the short circuit current density was measured with the maximum value obtained for the sample with nanoparticles on the native SiO_2.

Figure 6.20 shows the changes in absorption evidencing that surface plasmons can increase the spectral response over almost the entire solar spectrum.

[106]Stuart HR, Hall DG (1996) Appl Phys Lett. 69:2327; Stuart HR, Hall DG (1998) Appl Phys Lett. 73:3815; Schaadt DM et al (2005) Appl. Phys. Lett. 86:063106; Derkacs D et al (2006) Appl Phys Lett. 89:093103; Matheu P et al (2008) Appl Phys Lett. 93:113108; Pillai et al (2007) J Appl Phys. 101:093105; Derkacs D et al (2008) Appl Phys Lett. 93:091107; Nakayama et al (2008) Appl Phys Lett. 93:121904

[107]Pillai S et al (2007) J Appl Phys. 101:093105

[108]Beck FJ et al (2009) J Appl Phys. 105:114310

[109]Park J et al (2014) Appl Phys Lett. 104:033903

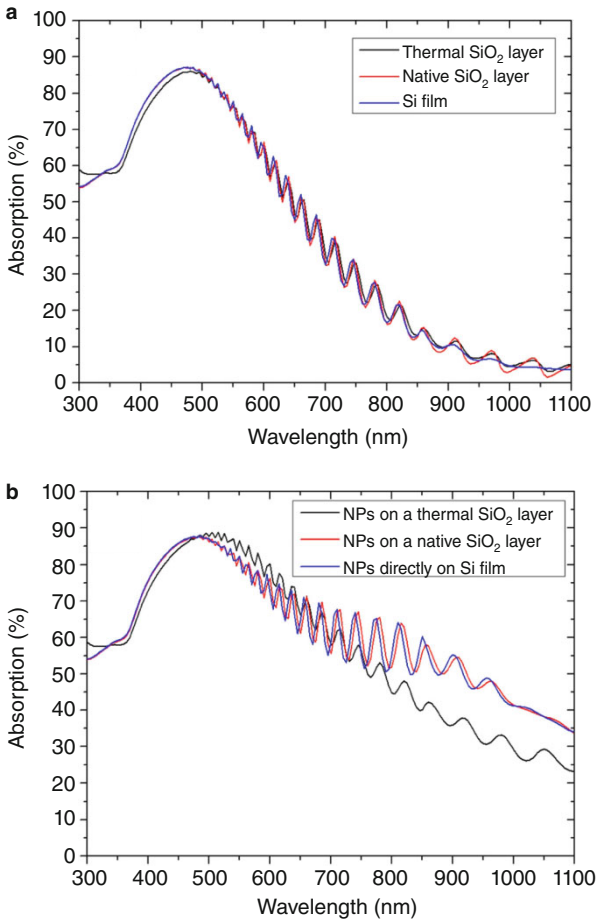


Fig. 6.20 Absorption of the three different surface conditions. (**a**) before and (**b**) after nanoparticles (NPs) formation. (From J Park et al., Appl. Phys. Lett. 104, 033903 (2014) reproduced with permission AIP)

An alternative use of resonant plasmon excitation in thin-film solar cells is to take advantage of the strong local field enhancement around the metal nanoparticle to increase absorption in a surrounding semiconductor material. The nanoparticles act as an effective antenna for the incident sunlight and the plasmonic near-field is coupled to the semiconductor increasing its effective absorption cross-section.[110]

[110]Rande B et al (2004) J Appl Phys. 96:7519; Kim SS et al (2008) Appl Phys Lett. 93:073307; Morfa AJ et al (2008) Appl Phys Lett. 92:013504; Lindquist NC et al (2008) Appl Phys Lett. 93:123308; Kume T et al (1995) Jpn J Appl Phys. 34:6448; Westphalen M et al (2000) Sol Energy Mater Sol. C61:97; Hagglund C et al (2008) Appl Phys Lett. 92:013113; Konda RB et al (2007) Appl Phys Lett. 91:191111; Hagglund C et al (2008) Appl Phys Lett. 92:053110; Kirkengena M et al (2007) J Appl Phys. 102:093713

A corrugated metallic film on the back surface of thin photovoltaic absorber layer can couple sunlight into SPP modes supported at the metal/semiconductor interface as well as guided modes in the semiconductor slab, whereupon the light is converted to photocarriers in the semiconductor. SPPs can be excited by the sunlight at the metal semiconductor interface and trap and guide light in the semiconductor layer.[111]

All these light-trapping techniques may allow considerable shrinkage of the photovoltaic layer thickness, while keeping the optical absorption (and thus efficiency) constant.

Other structures can be used. General overviews of the applications of plasmonic to this field have been made.[112]

6.8 Plasmon Microscopy

A scanning near-field microscope (SNOM) can allow observations with a nanoscale resolution. However the probe that is used as emitter, collector or scatterer of light is a small hole formed on the end of a metal-coated tapered optical fibre, and the efficiency is limited by the small amount of light that can be coupled through the hole. SPPs can be used to increase the efficiency.

Figure 6.21 shows a first realization by Sprecht et al.[113] in which a lateral resolution of 3 nm was achieved. A silver film was evaporated onto a glass prism. An incident laser beam was coupled to a plasmon mode at the silver air interface via the prism in the Kretschmann configuration. The excitation of the plasmon is recognized as a minimum in the reflected light intensity. A tungsten stylus with a radius of curvature of about 10 nm is placed close to the surface of the object and scatters the plasmons. The interaction depends strongly on the distance of the stylus from the surface. By raster scanning the tip over the sample surface, nanometer-scale maps of the interaction strength related to the topography of the object (in this case the silver film) can be recorded as a variation of the intensity of the reflected beam.

The system was later implemented by Ketterson.[114]

It is also possible to use a nanoparticle, much smaller than the light wavelength, to scan the sample surface. This system involves the use of a confined light field

[111]Giannini V et al (2008) Opt Express 16:19674; Mapel JK et al (2007) Appl Phys Lett. 90:121102; Tvingstedt K et al (2007) Appl Phys Lett. 91:113514; Heidel TD et al (2007) Appl Phys Lett. 91:093506

[112]Atwater HA, Polman A (2010) Nature Mat. 9:205; Mandal P, Sharma S (2016) Renewable and Sustainable Energy Reviews 65:537; Erwin WR et al (2016) Energy Environ Sci. 9:1577

[113]Sprecht M et al (1992) Phys Rev Lett. 68:476

[114]Kim YK et al (1995) Appl Phys Lett. 66:3407; Kryukov AE et al (1997) J. Appl. Phys. 82:5411

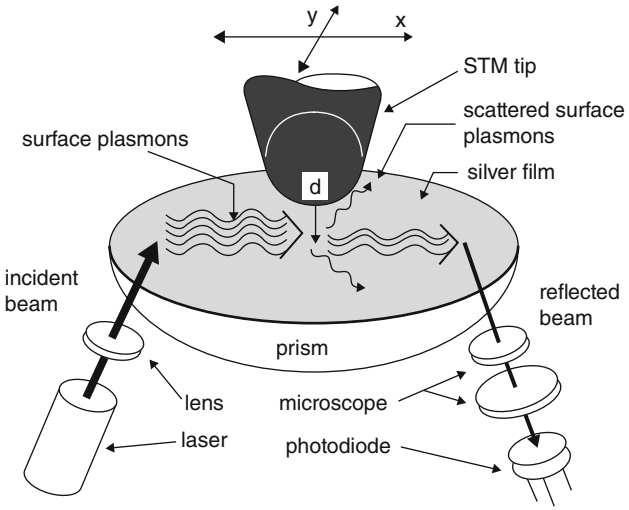


Fig. 6.21 Scheme of the set-up for a Plasmon microscope. Surface plasmons, excited by an incident laser beam, interact with a thin tungsten tip positioned very close to a silver surface. (From M Sprecht et al., Phys. Rev. Lett. 68, 476 (1992) reproduced with permission from AIP)

which then couples with propagating photons and allows to have a spatial resolution comparable to the size of the confined field, which is the same as the size of the metal nanostructure.[115]

A surface Plasmon microscopy to study the spatiotemporal patterns in electrochemical systems was demonstrated (Fig. 6.22) using the potential dependence of resonance conditions for the excitation of surface plasmons.[116]

If SPPs are used to illuminate interfacial structures of low-contrast samples, high contrast without loss of spatial resolution can be obtained owing to the high sensitivity of the plasmon resonance coupling to (for example) small optical thickness variations of thin dielectric coatings.[117]

6.9 Black-body Spatial and Temporal Coherence

The black-body radiation is usually regarded as an incoherent phenomenon, exhibiting a (transverse) spatial correlation, at a given wavelength λ, over a distance of

[115]Achermann M et al (2007) Opt Lett. 32:2254, ce l'ho

[116]Flatgen G et al (1995) Science 269:668

[117]Rothenhausler B, Knoll W (1988) Nature 332:615

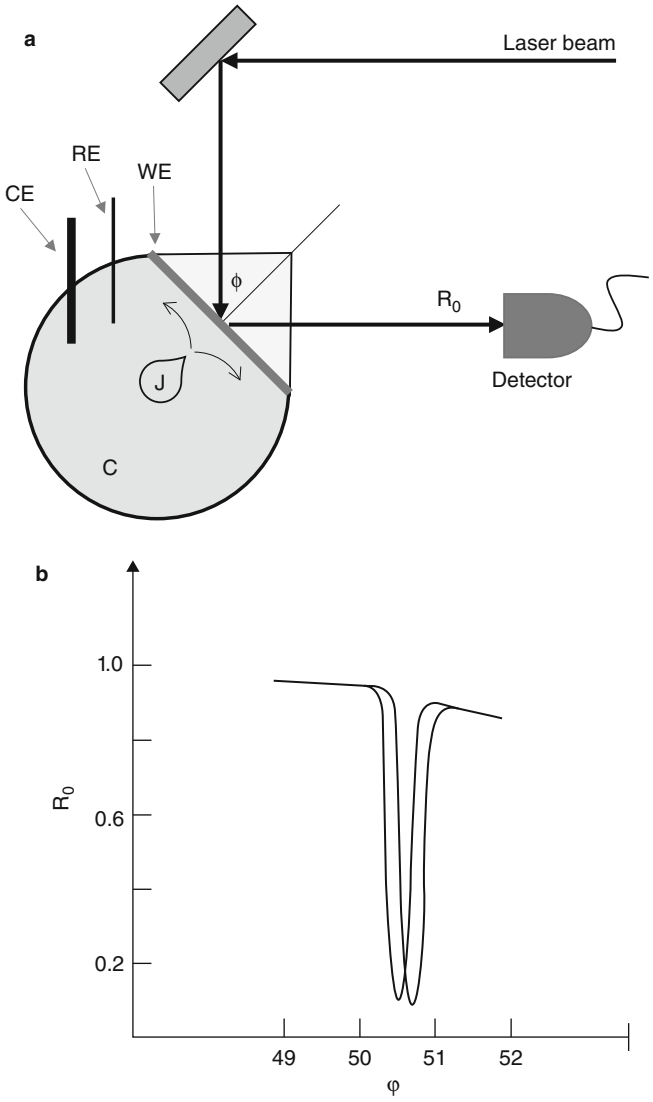


Fig. 6.22 (**a**) Experimental set-up. WE is the working electrode, RE the reference electrode, CE the counter electrode, J the impinging jet and C is the electrochemical cell. (**b**) Reflected intensity Ro versus the angle of incidence at two different values of the potentil drop across the electrode-electrolyte interface. (Modified from G Flatgen et al. Science 269, 668 (1995))

the order of $\lambda/2$. In deriving this result, the non-radiating near-field part of the emitted light is ignored, because it plays no role in the far field properties of emission from planar sources. Nevertheless, there are cases in which near-fields are detected, as for example in local-probe thermal microscopy.

The spatial correlation properties of the field emitted by a body are described by the correlation function in the space-frequency domain[118]

$$W(r_1, r_2, \omega) = <E(r_1\omega)E^*(r_2\omega)> \tag{6.12}$$

where E(r,ω) is the time-domain Fourier transform of the electric field E(r,t). To calculate this function we introduce the time-domain Fourier transform j(r,ω) of the random current density j(r,t). Considering a flat homogeneous opaque material filling the half-space $z < 0$, in local thermodynamic equilibrium at a uniform temperature T, the electric field at a given point r in the half space $z > 0$ is given by

$$E(r, \omega) = i\mu_o\omega \int G(r, r', \omega) \cdot j(r', \omega)d^3r' \tag{6.13}$$

where the integration is performed on the volume V of the body. G(r,r′,ω) is the Green dyadic of the vector Helmoltz equation describing the flat interface separating the medium from vacuum.

The spatial correlation function of the thermal current fluctuations is given by the fluctuation-dissipation theorem

$$<j(r, \omega)j(r', \omega)> = (\omega/\pi)\varepsilon_o\varepsilon''(\omega)\Theta(\omega, T)\delta(r - r')\delta(\omega - \omega') \tag{6.14}$$

where

$$\Theta(\omega, T) = \hbar\omega/2 + \hbar\omega/[\exp(\hbar\omega/kT) - 1] \tag{6.15}$$

is the mean energy of the quantum harmonic oscillator in thermal equilibrium at T. The quantity $\varepsilon''(\omega)$ is the imaginary part of the dielectric constant of the medium.

Inserting (6.13) into (6.12) and using (6.14), we obtain the cross-spectral function

$$W(\mathbf{r}_1, \mathbf{r}_2, \omega) = (\omega^3/\pi)\mu_o^2\varepsilon_o\varepsilon''\Theta(\omega, T) \int G(\mathbf{r}_1, \mathbf{r}', \omega)G^*(\mathbf{r}_2, \mathbf{r}', \omega)d^3\mathbf{r}' \tag{6.16}$$

with

$$G(\mathbf{r}, \mathbf{r}', \omega) = (i/8\pi^2) \int (1/\gamma_2)(\mathbf{s}t_s\mathbf{s} + \mathbf{p}_1t_p\mathbf{p}_2)\exp i\left[\mathbf{K} \cdot (\mathbf{R} - \mathbf{R}') + (\gamma_1 z - \gamma_2 z')\right]d^3\mathbf{K} \tag{6.17}$$

where

$$\mathbf{r} = (\mathbf{R}, z), \mathbf{s} = (\mathbf{K}/K) \times \mathbf{u}_z, \quad \mathbf{p}_j = (|K|\mathbf{u}_z + \gamma_j\mathbf{K}/K)/k_j, \quad k_1 = k = \omega/c, \quad k_2 = (\varepsilon)^{1/2}k$$

[118] We follow the calculations by Carminati R, Greffet JJ (1999) Phys Rev Lett. 82:1660

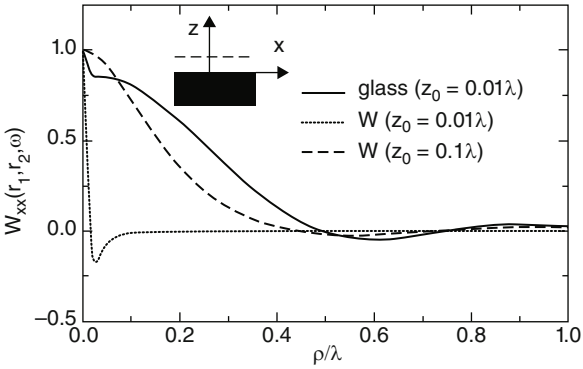


Fig. 6.23 Correlation function $W(r_1,r_2,\omega)$ versus ρ in planes of different z values. The solid black curve is for glass at z = 0.01λ.The two other curves are for tungsten at z = 0.01 λ and z = 0.1λ. All curves are normalized by their maximum value at ρ = 0. (From R Carminati and JJ Greffet, Phys. Rev. Lett. 82, 1660 (1999) reproduced with permission AIP)

and

$$\gamma_j = (k_j^2 - K^2)^{1/2}$$

being Re(γj)>0 and Im(γj)>0 and $t_s(K)$ and $t_p(K)$ are the Fresnel transmission factors for s and p polarizations, respectively.

Expression (6.16) has been solved numerically by Carminati and Greffet.[119] Figure 6.23 shows the spatial correlation as a function of $\boldsymbol{\rho} = |\mathbf{r}_1 - \mathbf{r}_2|$ at various distances from the medium surface for glass and tungsten at the wavelength λ = 500 nm.

In the very near field (z = 0.01λ), the curve corresponding to glass drops to negligible values after ρ = λ/2, showing that the correlation length is λ/2. In tungsten, on the contrary, the correlation length is much smaller than λ/2 on the order of 0.06λ (a distance comparable with the skin depth of tungsten at 500 nm) and at larger distances (z = 0.1λ) the correlation length is of the order of λ/2. This sub-wavelength correlation length is a pure near-field effect, due to non-radiative evanescent fields.

The behaviour is totally different in materials supporting surface plasmons or surface phonon polaritons (Fig. 6.24).

Both gold and silver exhibit surface-plasmon resonances at 620 nm. Figure 6.24a clearly shows that whereas the spatial correlation length for tungsten is a fraction of the wavelength, the correlation length for gold and silver is much larger extending up to 16λ for gold and 65λ for silver (the figure stops at only 5λ). The same is shown (Fig. 6.24b) for SiC which exhibits a surface-phonon polariton resonance at λ = 11.36 μm and no resonance at λ = 9.1 μm.

These results explained previous experiments on SiC.[120]

[119]Carminati R, Greffet JJ (1999) Phys Rev Lett. 82:1660

[120]Hesketh PJ et al (1986) Nature 324:549; Hesketh PJ et al (1988) Phys Rev. B37:10803; Le Gall J et al (1997) Phys Rev. B55:10105

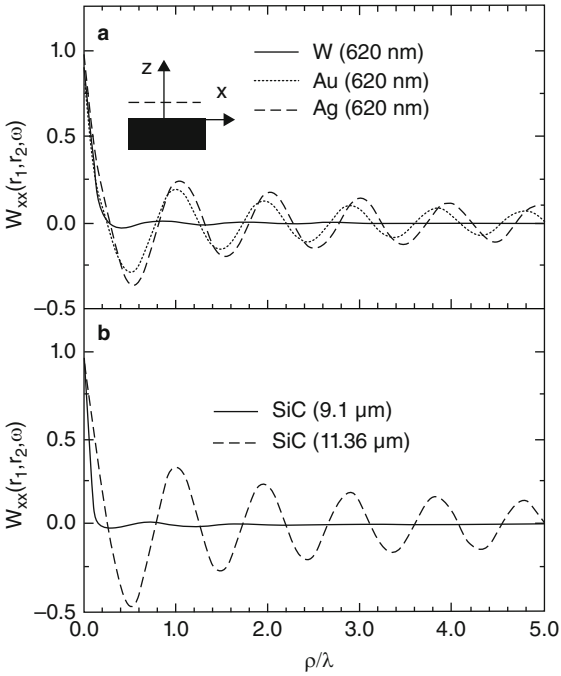


Fig. 6.24 Same as Fig. 6.10 with z = 0.05λ, (**a**) the case of three metals (tungsten, gold, and silver) at λ = 620 nm (**b**) the case of SiC at λ = 9.1 μm and λ = 11.36 μ. (From R Carminati and JJ Greffet, Phys. Rev. Lett. 82, 1660 (1999) reproduced with permission AIP)

Similar effects can be found in the near-field emission spectra (Fig. 6.25).[121]

The spectrum originates from the temporal coherence correlation function which can be derived with similar calculations as done for the spatial coherence and shows a marked influence of surface plasmons. The figure shows a dramatic difference between the black-body spectrum in the far-field (Fig. 6.25a) and near-field (Fig. 6.25c). The spectrum changes qualitatively on propagation due to surface waves existing in the region (150–180).10^{12} Hz due to the change in the local density of emitted electromagnetic modes as z changes.

In fact, the Planck spectrum I(ω) is obtained multiplying the thermal energy

$$\Theta(\omega, T) = \hbar\omega/[\exp(\hbar\omega/k_B T) - 1]$$

of a quantum oscillator by the density of modes per unit volume

$$N(\omega)d\omega = \omega^2 d\omega/(\pi^2 c^3)$$

[121]Shchegrov AV, Joulain K, Carminati R, Greffet JJ (2000) Phys RevLett. 85:1548

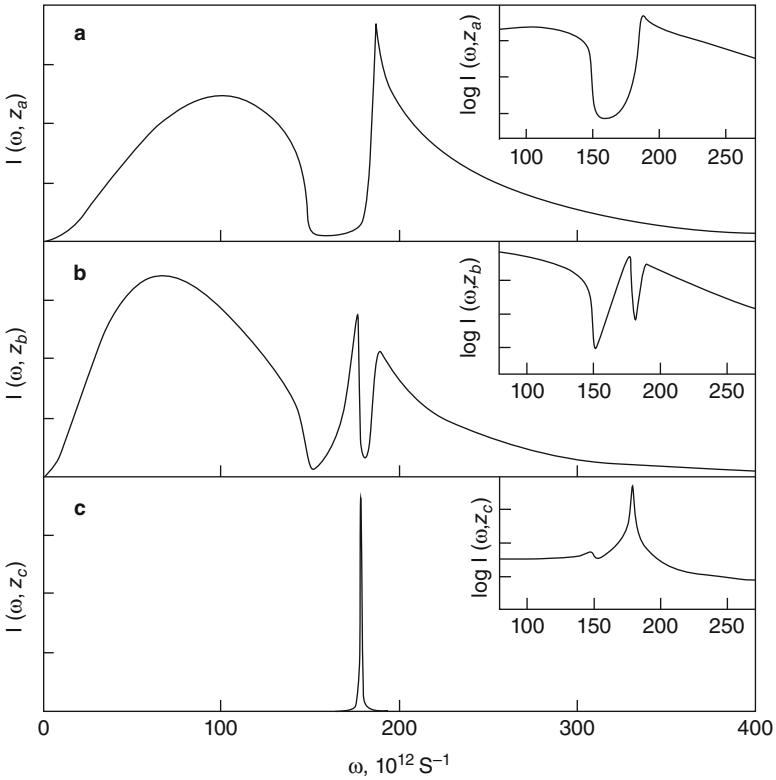


Fig. 6.25 Spectra of thermal emission of a semi-infinite sample of SiC at T = 300 K at three different heights above the surface: (**a**) z = 1000 μm, (**b**) z = 2 μm, and (**c**) z = 0.1 μm. The insert magnify the spectra plotted on a semilog scale in the range of strong contribution from evanescent surface modes. (From A V Shchegrov et al., Phys. Rev. Lett. 85, 1548 (2000) reproduced with permission from AIP)

in the frequency interval (ω,ω + dω), and dividing the result by dω

$$I(\omega) = \Theta(\omega, T)N(\omega) = (\hbar/\pi^2 c^3)\omega^3/[\exp(\hbar\omega/k_B T) - 1]$$

where T is the body temperature and k_B is Boltzmann's constant. The density of modes depends on z, and can be calculated by the dispersion relation of surface plasmons.

Related to these findings, another fundamental quantity is the degree of polarization that shows strongly polarization of the black-body radiation emitted when resonant surface plasmons or polaritons are present.[122]

[122]Setala T, Kaivola M, Friberg A (2002) Phys Rev Lett. 88:123902

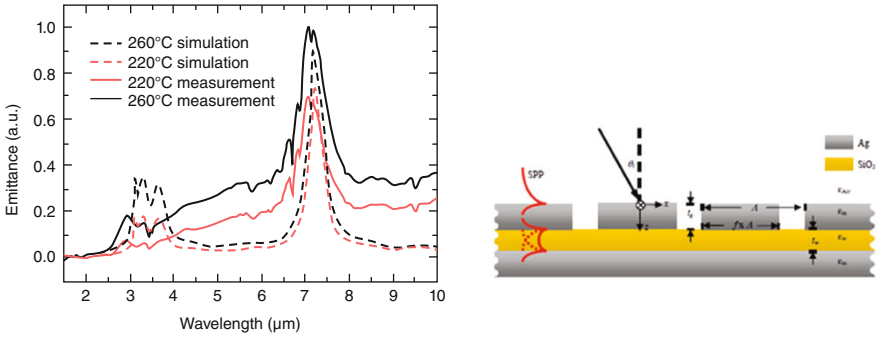


Fig. 6.26 Simulated (dashed line) and measured (solid line) emission spectrum of the IR emitter for 220C (red line) and 260C (black line), respectively. The grating (shown in the inset) had a thickness of 100 nm and a periodicity of 3000 nm. (From C M Wang et al., Opt. Express 15, 14673 (2007) © Optical Society America)

6.10 Controlled Thermal Emission Using Plasma Resonances

Thermal emission control is based on Kirchhoff's law of thermal radiation, which states that the emissivity of an object is equal to the absorptivity for a given frequency, direction, and polarization. Absorptivity depends on the magnitude of interaction between light and materials. One can increase absorptivity at a particular wavelength by using an optical resonance, which enhances the light-matter interaction at that wavelength. Surface plasmon polaritons can suit the scope.[123]

Special plasmonic structures can be used to control infrared thermal emission. For example Wang et al.[124] have built and studied a plasmonic multilayer structure which consisted of a relief metallic grating, a waveguide layer, and a metallic substrate. Due to the resonant coupling of plasmons by the grating the thermal emission spectrum can be strongly modified as shown in Fig. 6.26.

Similar results were obtained by patterned metal films.[125] The management of infrared thermal emission has been the subject of a number of studies.[126]

The use of surface-phonon polariton resonances in polar materials has also been exploited.[127]

[123]Ikeda K et al (2008) Appl Phys Lett. 92:021117; Miyazaki HT et al (2008) Appl Phys Lett. 92:141114; Biener G et al (2008) Appl. Phys. Lett. 92:081913

[124]Wang CM et al (2007) Opt Express 15: 14673

[125]Mason JA et al (2010) Opt. Express 18:25192

[126]Chang TH et al (2006) Appl Phys Lett. 89:173128; Tsai MW et al (2006) Appl Phys Lett. 89:173116; Kravets VG et al (2008) Phys Rev. B78:205405; Kreiter M et al (1999) Opt Comm. 168:117

[127]Greffet JJ et al (2002) Nature 416:61; Lee BJ et al (2005) Appl Phys Lett. 87:071904; Streyer W et al (2014) Appl Phys Lett. 104:131105

Chapter 7
Quantization of Evanescent Waves

7.1 Introduction

Macroscopic phenomena related to evanescent waves have been discussed in a classical framework in the preceding chapters. At the nanoscale, quantum phenomena become relevant, and a multiplicity of quanta or quasi particles appear on the scene.

The energy of the electromagnetic field is quantized, which means that it cannot take on a continuum of values, but rather energy values that are integer multiples of a specific amount of energy, or *quanta*, whose characteristics depend on the frequency, wavevector and polarization of the light. The energy quanta of the electromagnetic field are *photons*.

The energy of the oscillations of the free electron gas in a metal, or plasma of electrons, is also quantized. The energy quanta of the electron plasma oscillations in a bulk metal are the *plasmons*. At a dielectric/metal surface, electromagnetic waves can also excite surface plasmon waves, which involve a surface electron density wave and accompanying evanescent inhomogeneous waves in the dielectric and the metal. Their quanta of energy are the *surface plasmon*s, also called *surface plasmon polaritons* to emphasize the strong coupling between the electromagnetic field and the electron density oscillations.

In the quantization of the free electromagnetic field, i.e., the field propagating in the infinite vacuum in the absence of any charges or currents, it is customary to expand the field in homogeneous plane waves, and to admit no evanescent components. As should be evident by material in earlier chapters, this is not a valid procedure if sources, scatterers, apertures, etc., are present, and it is particularly not valid for the case of plasmonic devices, which involve current densities that support confined fields not representable by homogeneous plane waves.

Before tackling the quantization of evanescent waves, we present in the next two sections the standard CQED (Cavity Quantum Electrodynamics) quantization of the free electromagnetic field inside a unidimensional Fabry-Perot interferometer, which constitutes an example of canonical quantization for bosonic fields in the presence

© Springer International Publishing AG 2017

M. Bertolotti et al., *Evanescent Waves in Optics*, Springer Series in Optical Sciences,
DOI 10.1007/978-3-319-61261-4_7

of boundaries. Boundaries lead to a set of discrete classical modes, as opposed to the continuum of free space plane waves. We then proceed with the canonical quantization of a monochromatic electromagnetic field in a homogenous and lossless dielectric medium by means of a Fourier expansion in plane waves, followed by a quantization of the field within a finite volume. A short discussion on the properties of photon number and coherent states is followed by remarks on the concept of the photon and its non-localizability. These sections are intended to provide basic aspects of quantization of the electromagnetic field.

Evanescent waves, e.g., in the vicinity of vacuum/dielectric and dielectric/metal interfaces, are not included in a plane wave expansion. Since quantum fields should obey Maxwell's equations with the specific boundary conditions for a given problem, the initial step towards quantization in structured media is the expansion of the field in orthonormal modes that solve the classical problem.

In Section 7.7 we introduce a modal expansion valid for continuous or pulsed electromagnetic fields propagating through inhomogeneous or structured media, and discuss the appropriate mode normalization for quantization of the field energy. The normalization is valid if the optical properties of the media can be described by a space-dependent dielectric function.

Since evanescent waves originate in microscopic matter-light interactions, it might seem at first that to quantize them, a quantum mechanical description of the material and its interaction with the electromagnetic field is required. But modal expansions are an alternative to microscopic theories and can be used to quantize evanescent and surface plasmon waves. Carniglia and Mandel[1] were the first to introduce a modal expansion in orthonormalized evanescent waves as an alternative to the microscopic quantization procedure. The space is half filled with a homogeneous dielectric on one side of a flat surface and half empty on the other. The optical properties of the material and the vacuum are described by their passive refractive index, not altered by the electromagnetic field. Classical triplet modal functions that include not only the evanescent but the incident and reflected fields are obtained by solving the appropriate boundary problem at the interface. The TE and TM triplet modes for left and right incidence on the interface constitute an orthogonal basis for expansion of any arbitrary solution of the Maxwell equations, and quantization of evanescent waves in the absence of losses.

We then discuss the quantization of the free-electron gas oscillations in metals, or *bulk plasmons*, and finally we discuss the quantization of *surface plasmon polaritons*. The last are of great interest due to the possibility of achieving strong confinement of the electromagnetic field at scales far below that of conventional optics, where the quantum properties of the light become relevant and differ greatly from those of classical electromagnetic plane waves.

The last sections look at the results of recent research on *localized surface plasmons* (LSP), which can be excited in localized nanostructures, e.g., nanometal particles embedded in a dielectric. The corresponding fields are strongly localized

[1]Carniglia CK, Mandel L (1971) Phys Rev D3:280

around the nanostructure, and have been imaged by ultrafast electron microscopy[2]. Since the quantum properties of the LSP depend on the material and geometrical properties of the localized nanostructure (nano spheres, nano tubes, etc.), the outcome of light-matter interactions at the quantum level can be controlled and engineered, opening new possibilities for applications to single-photon sources and other quantum-controlled devices. A major step in this direction was the amplification of surface plasmon waves by stimulated emission, or *spaser*, that has allowed the realization of a spaser-based nano-scale laser. The study and applications of the quantum properties of light and its interaction with matter at the nanoscale constitutes the new rapidly expanding field of quantum plasmonics.[3] Advances in and prospective applications of this novel field are presented in the last section.

7.2 Quantization of the Electromagnetic Field in One Dimension

In this section, we present the basic concepts and procedures involved in the quantization of the electromagnetic field. The material presented here is intended for the reader not familiar with this subject and can be skipped by those familiar with quantum theory. We consider the simplest case, a one-dimensional Fabry-Perot resonator (Fig. 7.1a), which provides insight into the canonical quantization method with boundary conditions that will be used throughout the Chapter. It also introduces the reader to photon annihilation and creation operators, their commutation rules, and the properties of some specific quantum states of light.

Light is an electromagnetic wave, whose classical fields obey Maxwell's equations, supplemented with boundary conditions. In free space, boundary conditions can take the form of either vanishing or periodic fields at infinity. In the presence of matter — for example a dielectric or metal surface — the classical electromagnetic

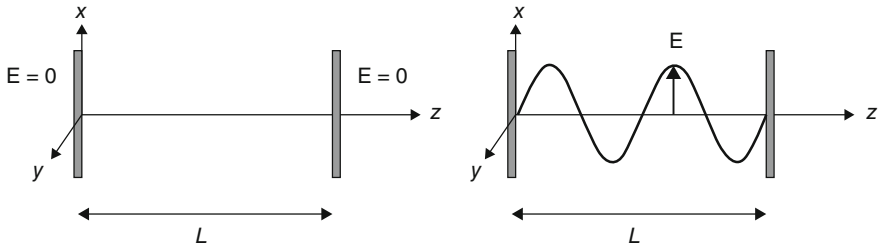


Fig. 7.1 (**a**) Fabry-Perot resonator consisting of two perfect mirrors separated by a distance L. We assume that the medium inside the resonator is the vacuum. (**b**) A mode $(m = 4)$ of the electric field inside the resonator. The corresponding magnetic field oscillation is in the *yz* – plane

[2]Piazza L et al (2015) Nature Comm. 6:6407

[3]Tame MS et al (2013) Nature Phys. 9:329

fields obey appropriate boundary conditions at the interfaces. Quantization of the electromagnetic field in the presence of boundaries thus usually starts with a solution to Maxwell's equations and the determination of the modes that satisfy the boundary conditions of the specific problem.

An ideal lossless one-dimensional Fabry-Perot resonator (Fig. 7.1a) consists of two flat, infinite and perfectly reflecting mirrors, separated by a distance L. Let us consider an x-polarized field $\mathrm{E}(z,t)$, which satisfies the wave equation:

$$\frac{\partial^2 \mathrm{E}}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 \mathrm{E}}{\partial t^2} = 0. \tag{7.1}$$

If the two mirrors are perfect conductors, the boundary conditions are $\mathrm{E}(z=0,\mathrm{L})=0$. We can find solutions corresponding to harmonic oscillating fields with frequency $\omega = ck$:

$$\mathrm{E}(z,t) = \mathrm{E}_0 e^{-i\omega t}\sin(kz) + c.c. \tag{7.2}$$

where $c.c.$ means complex conjugate. Boundary conditions require that k satisfy

$$k = \frac{m\pi}{\mathrm{L}}, \tag{7.3}$$

where L is the cavity length and m an integer. Different m values give rise to different modes of oscillation of the electromagnetic field inside the resonator (Fig. 7.1b).

We consider now only one mode, quantizing a mono-mode field characterized by wave-number k. Introducing a canonical position with harmonic temporal dependence

$$q(t) = \left(\frac{L\varepsilon_0}{2\omega^2}\right)^{\frac{1}{2}} \mathrm{E}_0 e^{-i\omega t} + c.c. \tag{7.4}$$

and its conjugate canonical momentum $p(t) = \dot{q}(t)$, we write the fields in the form

$$\mathrm{E}(z,t) = \left(\frac{2\omega^2}{L\varepsilon_0}\right)^{\frac{1}{2}} q(t)\sin(kz), \tag{7.5a}$$

$$\mathrm{B}(z,t) = \left(\frac{\mu_0\varepsilon_0}{k}\right)\left(\frac{2\omega^2}{L\varepsilon_0}\right)^{\frac{1}{2}} p(t)\cos(kz) \tag{7.5b}$$

The Hamiltonian of the field is obtained from the time averaged electromagnetic linear energy density as

$$H = \frac{1}{2}\int_0^L dz\left[\varepsilon_0 \mathrm{E}^2(z,t) + \frac{1}{\mu_0}\mathrm{B}^2(z,t)\right] = \frac{1}{2}\left(p^2 + \omega^2 q^2\right). \tag{7.6}$$

The Hamiltonian (7.6) has the same structure as the harmonic oscillator Hamiltonian for a particle of unit mass, but the oscillations of the canonical coordinate, instead of describing temporal variations of the position of a particle, describe the harmonic oscillation of a monochromatic electric field. So far, we have described the classical electromagnetic wave and the classical electromagnetic energy inside the resonator.

To quantize the energy of the field in energy quanta, or *photons*, we perform a canonical quantization, introducing the operators $\tilde{q}$ and $\tilde{p}$ for the canonical position and momentum respectively, and the commutation rule

$$[\tilde{q}, \tilde{p}] = \tilde{q}\tilde{p} - \tilde{p}\tilde{q} = i\hbar \tag{7.7}$$

where $\hbar$ is the Planck constant divided by 2π.

Canonical quantization is appropriate only for bosonic fields, i.e., for fields that obey Bose-Einstein statistics, which do not restrict the number of identical quanta occupying the same quantum state. Energy quanta were postulated by Einstein,[4] when he recognized that light emission and light-matter interactions were better understood if the light consisted of energy quanta, later to be called photons (a term introduced in a different context by G. N. Lewis[5]). Canonical quantization allows us to answer the question, "How many photons are there in this mode?," since we can introduce photon creation and annihilation operators $\tilde{a}^\dagger$ and $\tilde{a}$ respectively that add or remove one photon from the electromagnetic mode. They are defined as

$$\tilde{a}^\dagger = (2\hbar\omega)^{-\frac{1}{2}}(\omega\tilde{q} - i\tilde{p}) \tag{7.8a}$$

$$\tilde{a} = (2\hbar\omega)^{-\frac{1}{2}}(\omega\tilde{q} + i\tilde{p}) \tag{7.8b}$$

and satisfy the bosonic commutation relations

$$\left[\tilde{a}, \tilde{a}^\dagger\right] = \tilde{a}\tilde{a}^\dagger - \tilde{a}^\dagger\tilde{a} = 1 \tag{7.9a}$$

$$[\tilde{a}, \tilde{a}] = \left[\tilde{a}^\dagger, \tilde{a}^\dagger\right] = 0. \tag{7.9b}$$

The Hamiltonian for a single mode of the free electromagnetic field in the cavity can then be written as

$$\tilde{H} = \hbar\omega\left(\tilde{a}^\dagger\tilde{a} + \frac{1}{2}\right) = \hbar\omega\left(\tilde{n} + \frac{1}{2}\right) \tag{7.10}$$

[4]Einstein A (1905) Ann Phys. 17:132

[5]Lewis GN (1926) Nature 118:874

The Hermitian operator

$$\tilde{n} = \tilde{a}^{\dagger}\tilde{a}. \tag{7.11}$$

satisfies, in Dirac's notation, the eigenvalue equation

$$\tilde{n}|n\rangle = n|n\rangle. \tag{7.12}$$

It is called the *photon-number operator* because its eigenvalues n are positive integer numbers that correspond to a well-defined number of photons n in the quantum eigenstate $|n\rangle$. These quantum states $|n\rangle$ of the electromagnetic field with a well-defined number of quanta are called *number states* or *Fock states*, since they were first introduced and discussed by Vladimir Fock in his 1932 paper on second quantization[6] (see also[7]). The properties of the quantum field while in a Fock state will be discussed in the next section. Photon number states constitute a complete orthonormal basis set for the quantum states of the single-mode electromagnetic field:

$$\sum_{n=0}^{\infty} |n\rangle\langle n| = 1 \tag{7.13a}$$

$$\langle n|n'\rangle = \delta_{nn'}. \tag{7.13b}$$

The energy values of the single-mode Hamiltonian (7.10) can be written as:

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right), \quad n = 0, 1, 2, \ldots \tag{7.14}$$

The ground state is the vacuum state $|0\rangle$, a state in which there are no photons ($n = 0$) in the electromagnetic field. But even in the absence of photons there is an energy $\frac{\hbar\omega}{2}$ per mode of the electromagnetic field, known as the *zero point energy*, this energy being associated with electromagnetic field fluctuations that occur even in the vacuum state. The energy in a single mode of the field grows from the vacuum state zero point energy by integer multiples of $\hbar\omega$, which is the energy quantum, or the energy of a single *photon*, associated with the mode. The energy states are usually presented as the quantum states of a one-dimensional harmonic oscillator, as shown in Fig. 7.2.

We recall that the Hamiltonian (7.6) has the form of the Hamiltonian of a harmonic oscillator of unit mass, but the canonical position $q(t)$ describes the temporal variations of the electric field rather than any kind of mechanical oscillations.

[6]Fock V, Zeits F (1932) Phys. 75:622

[7]Faddeev LD et al (2004) Selected works. In: VA Fock Chapman & Hall, London

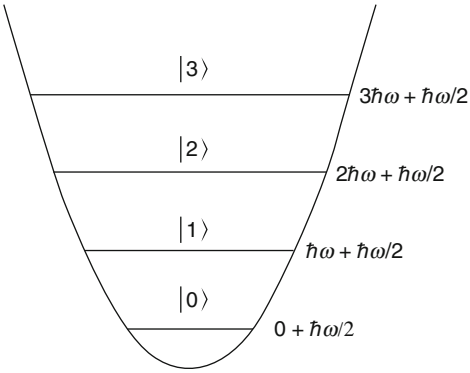


Fig. 7.2 Quantized energy states of a single-mode electromagnetic field. The levels inside the harmonic potential curve represent the energy of the first three photon number states. On the right are the corresponding energy values. The energy of consecutive photon number states differs by the energy of a single photon

7.3 Quantum States of the Electromagnetic Field

We want here to describe the quantum states of the electromagnetic field which are of interest in the present discussion.

7.3.1 Fock States

Since the quantum mechanical description of the radiation field is about quantum excitations, the most basic quantum states of the electromagnetic field are those with a fixed number of photons in a particular mode. A single-mode of the electromagnetic field, when in a Fock state, contains a precise number of photons, whose frequency and momentum are determined by the frequency and wavevector of the field mode. Preparing and maintaining a pure number state in a cavity is extremely difficult,[8] requiring that the cavity losses for the specific mode be very low and that additional sources of photon number fluctuations, such as those due to thermal fields, be eliminated. Superconducting cavities with quality factors as high as 3×10^{10}, kept at temperatures below 300 mK, are used for this purpose. Pioneer work on non-destructive counting of photons was done by H. Walther and S. Haroche, Nobel Prize 2012, who in 2007 stored microwave photons in a superconducting cavity for up to half a second and recorded the birth and death of a single-photon in the cavity.[9]

In a mathematical sense Fock states can be generated by applying the photon creation operator (7.8a) n times to the quantum vacuum state $|0\rangle$.

It can be easily shown that aside from a normalization factor, the state $\tilde{a}^\dagger|n\rangle$ is a new eigenstate of the photon number operator with eigenvalue $n+1$:

$$\tilde{n}\tilde{a}^\dagger|n\rangle = \tilde{a}^\dagger\tilde{a}\tilde{a}^\dagger|n\rangle = \tilde{a}^\dagger\left(\tilde{a}^\dagger\tilde{a}+1\right)|n\rangle = (n+1)\tilde{a}^\dagger|n\rangle. \tag{7.15}$$

[8] Varcoe BTH et al (2000) Nature 403:743

[9] Gleyzes S et al (2007) Nature 446:297

Therefore we can write $\tilde{a}^\dagger|n\rangle = c_n|n+1\rangle$ and determine c_n by calculating

$$\langle n|\,\tilde{a}\tilde{a}^\dagger|n\rangle = |c_n|^2\langle n+1|n+1\rangle = |c_n|^2 = \langle n|\tilde{a}^\dagger\tilde{a}+1|n\rangle = n+1, \tag{7.16}$$

yielding

$$\tilde{a}^\dagger|n\rangle = \sqrt{n+1}\,|n+1\rangle \tag{7.17}$$

Analogous considerations lead to

$$\tilde{a}|n\rangle = \sqrt{n}\,|n-1\rangle, \tag{7.18}$$

with $\tilde{a}|0\rangle = 0$. By using (7.17), we can build the entire ladder of photon number states from the vacuum:

$$|n\rangle = \frac{1}{\sqrt{n!}}\tilde{a}^{\dagger n}|0\rangle. \tag{7.19}$$

Until now the quantization formalism has not shown an explicit dependence on time. However, that dependence is implicit in the creation and annihilation operators, whose equations of motion (in the Heisenberg picture) are

$$\frac{d\tilde{a}(t)}{dt} = \frac{i}{\hbar}[H,\tilde{a}(t)] = \frac{i}{\hbar}\left[\hbar\omega\tilde{a}^\dagger(t)\tilde{a}(t),\tilde{a}(t)\right] = -i\omega\tilde{a}(t) \tag{7.20a}$$

$$\frac{d\tilde{a}^\dagger(t)}{dt} = \frac{i}{\hbar}\left[H,\tilde{a}^\dagger(t)\right] = \frac{i}{\hbar}\left[\hbar\omega\tilde{a}^\dagger(t)\tilde{a}(t),\tilde{a}^\dagger(t)\right] = i\omega\tilde{a}^\dagger(t). \tag{7.20b}$$

The creation and annihilation operators have a harmonic temporal dependence:

$$\tilde{a}^\dagger(t) = \tilde{a}^\dagger(0)e^{-i\omega t} \equiv \tilde{a}^\dagger e^{-i\omega t} \tag{7.21a}$$

$$\tilde{a}(t) = \tilde{a}(0)e^{i\omega t} \equiv \tilde{a}e^{i\omega t}. \tag{7.21b}$$

By performing the canonical quantization we are then able to answer the question posed earlier, "How many photons are there in a given mode of the EM field?" and to describe the temporal evolution of the creation and annihilation operators. But what information do we have on the electric field, which in classical electromagnetism fully characterizes the electromagnetic field?

Since the classical canonical position q has become the quantum operator $\tilde{q}$, the electric field has also become an operator. Combining (7.21a and b) in (7.8a and b), we obtain the electric field operator

$$\tilde{\mathrm{E}}(z,t) = \left(\frac{\hbar\omega}{L\varepsilon_0}\right)^{1/2}\left[\tilde{a}^\dagger e^{i\omega t} + \tilde{a}e^{-i\omega t}\right]\sin(kz). \tag{7.22}$$

Operators do not have a value on their own. If they correspond to an observable physical quantity like the electric field, they have an expected value, or quantum mechanical ensemble average, in each state of a quantum system. Note that the electric field operator consists of a linear combination of creation and annihilation operators. When these operators act on a Fock state, they produce a new Fock state that differs from the original one. Since Fock states are orthonormal, the quantum mechanical expected value of the field operator in a Fock state is always zero, irrespective of how many photons there are in the field mode:

$$\langle n|\tilde{\mathrm{E}}|n\rangle = 0. \tag{7.23}$$

That means that the phase of the electric field in a Fock state is completely undefined, and can take randomly any value between 0 and 2π. The expected value of the electric field is identically zero everywhere in the cavity and does not look at all like the classical wave depicted in Fig. 7.1. On the other hand, the expected value of the square of the electric field,

$$\langle n|\tilde{\mathrm{E}}^2|n\rangle = \frac{2\hbar\omega}{L\varepsilon_0}\left(n+\frac{1}{2}\right)\sin^2 kz, \tag{7.24}$$

increases in proportion to the number of photons in the mode, since it is proportional to the electric energy density in the cavity. Furthermore, the quantum fluctuations of the electric field, defined as

$$\Delta E = \left[\langle n|\tilde{\mathrm{E}}^2|n\rangle - \langle n|\tilde{\mathrm{E}}|n\rangle^2\right]^{\frac{1}{2}} = \left[\frac{2\hbar\omega}{L\varepsilon_0}\left(n+\frac{1}{2}\right)\right]^{1/2}\sin kz \tag{7.25}$$

increase as $\sqrt{n+1/2}$, which means that the electric field uncertainty increases with the number of photons in the mode. The state with minimum uncertainty in the value of the electric field is the vacuum, for which

$$\Delta \mathrm{E}_{vac} = \sqrt{\frac{\hbar\omega}{L\varepsilon_0}}\sin kz. \tag{7.26}$$

7.3.2 Coherent States

The next question is, "What type of quantum state describes a field that better resembles the classical electromagnetic field in the cavity, i.e., a quantum state with minimum uncertainty in the electric field?" Such states are called coherent states and are eigenstates of the annihilation operator:

$$\tilde{a}|\alpha\rangle = \alpha|\alpha\rangle. \tag{7.27}$$

The expected value of the electric field operator becomes

$$\langle\alpha|\tilde{\mathrm{E}}(z,t)|\alpha\rangle = \left(\frac{\hbar\omega}{L\varepsilon_0}\right)^{\frac{1}{2}}\left[\alpha^* e^{i\omega t} + \alpha e^{-i\omega t}\right]\sin(kz), \tag{7.28}$$

which resembles the classical wave except for the fact that its variance is not zero but minimal and coincides with the variance of the electric field in the vacuum state of the field.

Coherent states can be expressed as a specific *wave packet*, a superposition of Fock states, that gives rise to a minimum uncertainty:

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle. \tag{7.29}$$

Coherent states have minimum uncertainty in amplitude and phase, but the exact number of photons in the wave packet is not known. From (7.29) we can find the expected value of the photon number operator in a coherent state,

$$\langle\alpha|\tilde{n}|\alpha\rangle = |\alpha|^2, \tag{7.30}$$

and the probability of finding n photons, also known as *photon statistics*,

$$P(n) = \frac{\alpha^{2n}}{n!}e^{-|\alpha|^2}, \tag{7.31}$$

which is a Poisson distribution with parameter $|\alpha|^2$. From the properties of Poisson distributions, we know that the average number of photons is equal to the variance of the distribution; therefore,

$$\Delta n = \left[\langle n^2\rangle - \langle n\rangle^2\right]^{\frac{1}{2}} = \sqrt{\langle n\rangle} = |\alpha|. \tag{7.32}$$

The ratio $\Delta n/\langle n\rangle$ scales as $1/\sqrt{\langle n\rangle}$, which means that the fractional fluctuations diminish with the number of photons.

There exist other quantum states of the electromagnetic field with singular properties—the so-called squeezed states, for example—that will not be discussed here. (For a review of squeezed states of light and their applications to laser interferometers, including the improvement of the measurement sensitivity of GEO 600, a laser interferometer built for the detection of gravitational waves, see the online report by Schnabel.[10])

[10]Schnabel R (2016) https://arxiv.org/pdf/1611.03986.pdf [online]

The study of the properties of quantum states of the electromagnetic field and its interaction with matter is the object of quantum optics, a field that was born after the discovery of the laser, whose field is the closest realization of a quantum coherent field. R. Glauber is recognized as the father of quantum optics. He was awarded the Nobel Prize in Physics in 2005 "for his contribution to the quantum theory of optical coherence."[11]

7.3.3 The Multimodal Free Field

The previous sections were devoted to the quantization of the single mode field. Note that even in a one-dimensional cavity, there are in general many electromagnetic modes that satisfy the boundary conditions, characterized by the discrete set of wavenumbers of (7.3), and that are analogous to the modes of a vibrating string with fixed extremes.

In the limit when the cavity length tends to infinity, i.e., when the mirrors are separated by an infinite distance, we recover the propagation of the electromagnetic field in free-space. The wavenumber k (7.3) becomes a continuous variable, corresponding to an infinite number of modes of different frequencies propagating in the z-direction (one-dimensional propagation in free-space).

In the next section, we quantize the three-dimensional electromagnetic field in lossless, non-magnetic, homogenous dielectric media, whose relative permittivity ε is real. The results apply to the quantization of the free field in vacuum, when ε is replaced by the permittivity of vacuum ε_0.

7.4 Quantization of the Electromagnetic Field in Lossless Dielectric Media

7.4.1 Maxwell's Equations and the Reciprocal Space

Lenses, optical fibers, and various other optical components are made of dielectric materials that are globally neutral (do not have free charges) and whose electron clouds are tightly bound and localized around the ions: space charges and currents do not exist. Dielectric media are described by Equations (1.1) with $\rho_{ext} = 0$ and $\mathbf{j}_{ext} = 0$. The interaction of the free-space light field with the dielectric, within the dipole approximation, results in forced oscillations of the highly-localized electron clouds, inducing oscillating atomic dipoles. Therefore, dielectric media respond to light by becoming polarized and/or magnetized.

[11] Glauber RJ (2007) Quantum theory of optical coherence. Selected papers and lectures (A collection of reprints of Glauber's most important papers from 1963 to 1999, selected by the author). Weinheim, Wiley, VCH

The magnetic field **B** can be expressed in terms of a vector potential **A** as

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{7.33}$$

From (1.1c), the electric field **E** then becomes

$$\mathbf{E} = -\dot{\mathbf{A}}. \tag{7.34}$$

The vector potential is not defined unequivocally. Adding to it the gradient of any scalar field leaves the **B** field in (7.33) unmodified. It is customary in quantum optics to work within the Coulomb gauge, also called the transverse gauge, which requires that

$$\nabla \cdot \varepsilon \mathbf{A} = 0. \tag{7.35}$$

Applying the curl operation to (7.33) we obtain a wave equation for the vector potential:

$$\frac{1}{\varepsilon}\nabla \times \frac{1}{\mu}\nabla \times \mathbf{A} - \frac{1}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = 0. \tag{7.36}$$

In a bulk isotropic dielectric with uniform ε, μ, and refractive index n, the wave equation for the vector potential is like those for the **E** and **B** fields:

$$\nabla^2 \mathbf{A} - \frac{n^2}{c^2}\frac{\partial^2 \mathbf{A}}{\partial t^2} = 0. \tag{7.37}$$

To quantize the electromagnetic field, we follow C. Cohen-Tannoudji et al.,[12] and first write the Maxwell equations and the fields in the reciprocal space or **k** space. The electric field $\mathcal{E}(\mathbf{k}, t)$ in the reciprocal space is related to $\mathbf{E}(\mathbf{r}, t)$ by the Fourier transform operation and its inverse:

$$\mathcal{E}(\mathbf{k}, t) = \frac{1}{(2\pi)^{3/2}} \int d\mathbf{r}\, \mathbf{E}(\mathbf{r}, t) e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{7.38a}$$

$$\mathbf{E}(\mathbf{r}, t) = \frac{1}{(2\pi)^{3/2}} \int d\mathbf{k}\, \mathcal{E}(\mathbf{k}, t) e^{-i\mathbf{k}\cdot\mathbf{r}}. \tag{7.38b}$$

Since $\mathbf{E}(\mathbf{r}, t)$ is real, from (7.38a) it follows that

$$\mathcal{E}^*(\mathbf{k}, t) = \mathcal{E}(-\mathbf{k}, t). \tag{7.39}$$

[12] Cohen-Tannoudji C et al (1998) Photons and atoms: introduction to quantum electrodynamics. John Wiley & Sons.

Analogous equations relate the fields $\mathbf{B}(\mathbf{r},t)$ and $\mathbf{A}(\mathbf{r},t)$ with the corresponding Fourier transforms $\mathcal{B}(\mathbf{k},t)$ and $\mathcal{A}(\mathbf{k},t)$. Working in the reciprocal space has the advantage that the gradient operator ∇ in real space transforms into multiplication by $i\mathbf{k}$ in the reciprocal space, and Maxwell's equations become

$$i\mathbf{k}\cdot\varepsilon\mathcal{E}=0 \tag{7.40a}$$

$$i\mathbf{k}\cdot\mathcal{B}=0 \tag{7.40b}$$

$$i\mathbf{k}\times\mathcal{E}=-\dot{\mathcal{B}} \tag{7.40c}$$

$$i\mathbf{k}\times\mathcal{B}=-\mu\varepsilon\dot{\mathcal{E}}, \tag{7.40d}$$

where the dot means partial derivative with time. In a lossless dielectric the magnitude of the wave vector $\mathbf{k}$ is $\mathrm{k}=\omega_{\mathrm{k}}\sqrt{\varepsilon\mu}$.

From (7.40a) and (7.40b) we see that in a homogenous dielectric the fields are transverse.

Using the Parseval identity (Rayleigh's theorem) we can write the electromagnetic field energy as

$$H=\frac{1}{2}\int d\mathbf{r}\left[\varepsilon\mathbf{E}^2(\mathbf{r},t)+\frac{\mathbf{B}^2(\mathbf{r},t)}{\mu}\right]=\frac{1}{2}\int d\mathbf{k}\left[\varepsilon|\mathcal{E}(\mathbf{k},t)|^2+\frac{|\mathcal{B}(\mathbf{k},t)|^2}{\mu}\right]. \tag{7.41}$$

The relations between the fields and the vector potential in the reciprocal space are

$$\mathcal{B}=i\mathbf{k}\times\mathcal{A}, \tag{7.42a}$$

$$\mathcal{E}=-\dot{\mathcal{A}}. \tag{7.42b}$$

The linear momentum of the field is given by

$$\mathbf{P}=\frac{1}{\mu}\int d\mathbf{r}\,\mathbf{E}(\mathbf{r},t)\times\mathbf{B}(\mathbf{r},t)=\frac{1}{\mu}\int d\mathbf{k}\;\mathcal{E}^*(\mathbf{k},t)\times\mathcal{B}(\mathbf{k},t). \tag{7.43}$$

7.4.2 *Definition of normal variables*

Equations (7.40c) and (7.40d) are the equations of motion for the fields:

$$\mathbf{k}\times\dot{\mathcal{B}}=i\mathrm{k}^2\mathcal{E} \tag{7.44a}$$

$$\dot{\mathcal{E}}=\frac{i}{\mu\varepsilon}\mathbf{k}\times\mathcal{B}. \tag{7.44b}$$

The eigenfunctions for the system are

$$\frac{\partial}{\partial t}\left(\mathcal{E} \mp \frac{1}{\sqrt{\mu\varepsilon}}\hat{\mathbf{k}}\times\mathcal{B}\right) = \mp i\omega_{\mathrm{k}}\left(\mathcal{E} \mp \frac{1}{\sqrt{\mu\varepsilon}}\hat{\mathbf{k}}\times\mathcal{B}\right) \tag{7.45}$$

with

$$\omega_{\mathrm{k}} = \frac{\mathrm{k}}{\sqrt{\mu\varepsilon}}, \quad \hat{\mathbf{k}} = \frac{\mathbf{k}}{\mathrm{k}}. \tag{7.46}$$

We can define two new functions (Cohen-Tannoudji et al. 1998), or normal modes that are eigenfunctions of (7.44a and b):

$$\boldsymbol{\alpha}(\mathbf{k},t) = -\frac{i}{2\mathcal{N}(k)}\left(\mathcal{E} - \frac{1}{\sqrt{\mu\varepsilon}}\hat{\mathbf{k}}\times\mathcal{B}\right) \tag{7.47a}$$

$$\boldsymbol{\beta}(\mathbf{k},t) = -\frac{i}{2\mathcal{N}(k)}\left(\mathcal{E} + \frac{1}{\sqrt{\mu\varepsilon}}\hat{\mathbf{k}}\times\mathcal{B}\right), \tag{7.47b}$$

where $\mathcal{N}$ is a normalization factor that will be chosen later. $\boldsymbol{\alpha}(\mathbf{k},t)$ and $\mathcal{B}(\mathbf{k},t)$ are not independent. The real character of the electric and magnetic field requires that

$$\boldsymbol{\beta}(\mathbf{k},t) = -\boldsymbol{\alpha}^{*}(-\mathbf{k},t). \tag{7.48}$$

Inverting (7.47a and b) and using (7.48) leads to

$$\mathcal{E}(\mathbf{k},t) = i\mathcal{N}(k)\left[\boldsymbol{\alpha}(\mathbf{k},t) - \boldsymbol{\alpha}^{*}(-\mathbf{k},t)\right] \tag{7.49a}$$

$$\mathcal{B}(\mathbf{k},t) = i\mathcal{N}(k)\sqrt{\mu\varepsilon}\left[\hat{\mathbf{k}}\times\boldsymbol{\alpha}(\mathbf{k},t) + \hat{\mathbf{k}}\times\boldsymbol{\alpha}^{*}(-\mathbf{k},t)\right]. \tag{7.49b}$$

Note that $\mathcal{E}(\mathbf{k},t)$ and $\mathcal{B}(\mathbf{k},t)$ are determined from $\boldsymbol{\alpha}(\mathbf{k},t)$. Knowing $\boldsymbol{\alpha}(\mathbf{k},t)$ for all $\mathbf{k}$ allows us to determine the electric and magnetic fields in real space through the inverse Fourier transform.

The equation of motion for $\boldsymbol{\alpha}(\mathbf{k},t)$ is

$$\dot{\boldsymbol{\alpha}}(\mathbf{k},t) + i\omega_{\mathrm{k}}\boldsymbol{\alpha}(\mathbf{k},t) = 0. \tag{7.50}$$

Its temporal dependence corresponds to a harmonic oscillation with frequency ω_{k}.

Finally, we take into account the fact that for each vector $\mathbf{k}$ there are two possible polarizations. The vector $\boldsymbol{\alpha}(\mathbf{k},t)$ can then be expanded into two components with mutually orthogonal polarizations, each orthogonal to $\mathbf{k}$:

$$\boldsymbol{\alpha}(\mathbf{k},t) = \sum_{\sigma=1}^{2} \hat{\boldsymbol{e}}_{k\sigma}\alpha_{\sigma}(\mathbf{k},t).$$

7.4.3 Energy and Linear Momentum

Both classical and quantum fields carry energy and momentum. The canonical quantization of the electromagnetic field is performed after proper normalization of the classical field modes to assure that the electromagnetic energy in a mode, obtained by integration over the reciprocal (or direct) space, results in the energy of the photon associated with that mode. We proceed then to express the energy and momentum of the field in terms of $\boldsymbol{\alpha}(\mathbf{k},t)$, replacing (7.49a and b) into (7.41) and (7.43). The classical functions $\boldsymbol{\alpha}(\mathbf{k},t)$ and $\boldsymbol{\alpha}^*(-\mathbf{k},t)$ are numbers that commute, but in the quantization of the field, they become non-commuting operators. It is therefore important to respect the ordering of products of these two variables. After some algebra one obtains the following:

$$H=\varepsilon\int d\mathbf{k}\,\mathcal{N}^2(k)\left[\boldsymbol{\alpha}^*(\mathbf{k},t)\cdot\boldsymbol{\alpha}(\mathbf{k},t)+\boldsymbol{\alpha}(-\mathbf{k},t)\cdot\boldsymbol{\alpha}^*(-\mathbf{k},t)\right]. \tag{7.51}$$

Replacing $\mathbf{k}$ with $-\mathbf{k}$ in the second term of the integral allows the replacement of $\boldsymbol{\alpha}(-\mathbf{k},t)\cdot\boldsymbol{\alpha}^*(-\mathbf{k},t)$ by $\boldsymbol{\alpha}(\mathbf{k},t)\cdot\boldsymbol{\alpha}^*(\mathbf{k},t)$, and choosing $\mathcal{N}(k)=\sqrt{\hbar\omega_{\mathrm{k}}/2\varepsilon}$ leads to

$$H=\sum_{\sigma=1}^{2}\int d\mathbf{k}\frac{\hbar\omega_{\mathrm{k}}}{2}\left[\alpha_\sigma^*(\mathbf{k},t)\alpha_\sigma(\mathbf{k},t)+\alpha_\sigma(\mathbf{k},t)\alpha_\sigma^*(\mathbf{k},t)\right]. \tag{7.52}$$

This Hamiltonian is the sum of the energy of independent harmonic oscillators, each one associated with a wave vector $\mathbf{k}$ and a polarization σ, which together characterize a *mode* of the electromagnetic field.

A similar procedure regarding (7.43) leads to the following expression for the total linear momentum of the electromagnetic field:

$$\mathbf{P}=\sum_{\sigma=1}^{2}\int d\mathbf{k}\frac{\hbar\mathbf{k}}{2}\left[\alpha_\sigma^*(\mathbf{k},t)\alpha_\sigma(\mathbf{k},t)+\alpha_\sigma(\mathbf{k},t)\alpha_\sigma^*(\mathbf{k},t)\right]. \tag{7.53}$$

The electric and magnetic fields in real space can be obtained by the inverse Fourier transform:

$$\mathbf{E}(\mathbf{r},t)=\frac{i}{(2\pi)^{3/2}}\sum_{\sigma=1}^{2}\int d\mathbf{k}\,\mathcal{N}(k)\hat{\boldsymbol{e}}_{\mathbf{k}\sigma}\left[\alpha_\sigma(\mathbf{k},t)e^{-i\mathbf{k}\cdot\mathbf{r}}-\alpha_\sigma^*(\mathbf{k},t)e^{i\mathbf{k}\cdot\mathbf{r}}\right] \tag{7.54a}$$

$$\mathbf{B}(\mathbf{r},t)=\frac{i\sqrt{\mu\varepsilon}}{(2\pi)^{3/2}}\sum_{\sigma=1}^{2}\int d\mathbf{k}\,\mathcal{N}(k)\left(\hat{\mathbf{k}}\times\hat{\boldsymbol{e}}_{\mathbf{k}\sigma}\right)\left[\alpha_\sigma(\mathbf{k},t)e^{-i\mathbf{k}\cdot\mathbf{r}}-\alpha_\sigma^*(\mathbf{k},t)e^{i\mathbf{k}\cdot\mathbf{r}}\right]. \tag{7.54b}$$

7.4.4 Electromagnetic modes in a finite volume

Equation (7.54a and b) correspond to the expansion of the electromagnetic field in an infinite dielectric in plane waves. However, in many problems the electromagnetic field in a finite volume is required to satisfy specific boundary conditions. Gaussian modes of a laser cavity, the modes of an optical resonator, and LP modes in optical fibers are some examples. S. Haroche was awarded the Nobel Prize in 2012, jointly with D. Wineland, for developing cavity quantum electrodynamics (CQED),[13] a new field based on the study of radiation-matter interactions inside optical or microwave cavities that has enabled the control of radiative properties of atoms like enhancement or suppression of spontaneous emission, or the realization of quantum logic gates.

For the sake of simplicity, we consider the electromagnetic field contained in a cube of length L and volume $V = L^3$ and impose periodic boundary conditions at the cube walls that are satisfied by a discrete set of $\mathbf{k}$ vectors:

$$\mathbf{k}_{n,m,l} = \frac{2\pi}{L}(n\hat{\boldsymbol{x}} + l\hat{\boldsymbol{y}} + m\hat{z}), \tag{7.55}$$

with n, l, m being integer numbers. For each $\mathbf{k}$ vector there are two orthogonal polarizations $\mathbf{e}_{\mathbf{k}\sigma}$, and associated with each of them is a mode of the electromagnetic field. The integral over the reciprocal space is replaced by a sum over $\mathbf{k}$ obeying

$$\int d\mathbf{k} f(\mathbf{k}) \to \left(\frac{2\pi}{L}\right)^{\frac{3}{2}} \sum_j f(\mathbf{k}_j) \equiv \left(\frac{2\pi}{L}\right)^{\frac{3}{2}} \sum_{\mathbf{k}} f(\mathbf{k}). \tag{7.56}$$

The electric and magnetic field become

$$\mathbf{E}(\mathbf{r}, t) = i \sum_{\mathbf{k},\sigma} \sqrt{\frac{\hbar\omega_{\mathrm{k}}}{2\varepsilon V}}\, \hat{\boldsymbol{e}}_{\mathbf{k}\sigma} \left[\alpha_{\mathbf{k}\sigma} e^{i\mathbf{k}\cdot\mathbf{r}} - \alpha^*_{\mathbf{k}\sigma} e^{-i\mathbf{k}\cdot\mathbf{r}}\right] \tag{7.57a}$$

$$\mathbf{B}(\mathbf{r}, t) = i \sum_{\mathbf{k},\sigma} \sqrt{\frac{\hbar}{2\omega_{\mathrm{k}}\varepsilon V}} (\mathbf{k} \times \hat{\boldsymbol{e}}_{\mathbf{k}\sigma}) \left[\alpha_{\mathbf{k}\sigma} e^{i\mathbf{k}\cdot\mathbf{r}} - \alpha^*_{\mathbf{k}\sigma} e^{-i\mathbf{k}\cdot\mathbf{r}}\right], \tag{7.57b}$$

and the energy and the total linear momentum become:

$$H = \sum_{\mathbf{k},\sigma} \frac{\hbar\omega_{\mathrm{k}}}{2} \left[\alpha^*_{\mathbf{k}\sigma}\alpha_{\mathbf{k}\sigma} + \alpha_{\mathbf{k}\sigma}\alpha^*_{\mathbf{k}\sigma}\right] \tag{7.58a}$$

$$\mathbf{P} = \sum_{\mathbf{k},\sigma} \frac{\hbar\mathbf{k}}{2} \left[\alpha^*_{\mathbf{k}\sigma}\alpha_{\mathbf{k}\sigma} + \alpha_{\mathbf{k}\sigma}\alpha^*_{\mathbf{k}\sigma}\right]. \tag{7.58b}$$

[13] Walther H et al (2006) Rep Prog Phys. 69:1325

7.4.5 Heuristic Approach to the Quantization of the Electromagnetic Field

The classical formalism discussed in the previous sections leads the way to an elementary approach to quantization of the field. The normal variables are time dependent, therefore the quantization in the Heisenberg picture, where the quantum operators are time dependent, is straightforward. Since (7.58a and b) represent sums over the energy and momentum of independent harmonic oscillators, each associated with a mode $\mathbf{k}$, σ, we can replace the normal variables $\alpha_{\mathbf{k}\sigma}$ and $\alpha^*_{\mathbf{k}\sigma}$ by annihilation and creation operators $\tilde{a}_{\mathbf{k}\sigma}$ and $\tilde{a}^\dagger_{\mathbf{k}\sigma}$, respectively, which obey the commutation relations

$$\left[\tilde{a}_{\mathbf{k}\sigma}, \tilde{a}^\dagger_{\mathbf{k}'\sigma'}\right] = \delta_{\mathbf{k}\mathbf{k}'}\delta_{\sigma\sigma'} \tag{7.59}$$

$$\left[\tilde{a}^\dagger_{\mathbf{k}'\sigma'}, \tilde{a}^\dagger_{\mathbf{k}\sigma}\right] = \left[\tilde{a}_{\mathbf{k}'\sigma'}, \tilde{a}_{\mathbf{k}\sigma}\right] = 0. \tag{7.60}$$

The Hamiltonian and the linear momentum operators become

$$\tilde{H} = \frac{1}{2}\sum_{\mathbf{k},\sigma} \hbar\omega_{\mathrm{k}}\left(\tilde{a}^\dagger_{\mathbf{k}\sigma}\tilde{a}_{\mathbf{k}\sigma} + \tilde{a}_{\mathbf{k}\sigma}\tilde{a}^\dagger_{\mathbf{k}\sigma}\right) = \sum_{\mathbf{k},\sigma} \hbar\omega_{\mathrm{k}}\left(\tilde{n}_{\mathbf{k}\sigma} + \frac{1}{2}\right) \tag{7.61a}$$

$$\tilde{\mathbf{P}} = \sum_{\mathbf{k},\lambda} \hbar\mathbf{k}\, \tilde{a}^\dagger_{\mathbf{k}\sigma}\tilde{a}_{\mathbf{k}\sigma} = \sum_{\mathbf{k},\sigma} \hbar\mathbf{k}\, \tilde{n}_{\mathbf{k}\sigma}. \tag{7.61b}$$

The energy eigenstates of each independent oscillator are eigenstates of the photon number operators $\tilde{n}_{\mathbf{k}\sigma} = \tilde{a}^\dagger_{\mathbf{k}\sigma}\tilde{a}_{\mathbf{k}\sigma}$, and the total energy in the field is the sum of the energy of the photons in each mode plus the vacuum energy. The Fock states of the multimode field are the direct product of the Fock states of individual oscillators

$$\left|n_{\mathbf{k}_1\sigma_1}, n_{\mathbf{k}_2\sigma_2}, \ldots, n_{\mathbf{k}_j\sigma_j}, ..\right\rangle = \left|n_{\mathbf{k}_1\sigma_1}\right\rangle \otimes \left|n_{\mathbf{k}_2\sigma_2}\right\rangle, \ldots, \otimes \left|n_{\mathbf{k}_j\sigma_j}\right\rangle \otimes \ldots \tag{7.62}$$

where the symbol $\otimes$ means direct product. These states satisfy

$$\tilde{n}_{\mathbf{k}_j\sigma_j}\left|n_{\mathbf{k}_1\sigma_1}, n_{\mathbf{k}_2\sigma_2}, \ldots, n_{\mathbf{k}_j\sigma_j}, ..\right\rangle = n_{\mathbf{k}_j\sigma_j}\left|n_{\mathbf{k}_1\sigma_1}, n_{\mathbf{k}_2\sigma_2}, \ldots, n_{\mathbf{k}_j\sigma_j}, ..\right\rangle, \tag{7.63}$$

where $n_{\mathbf{k}_j\sigma_j}$ is the number of photons in the mode $\mathbf{k}_j\sigma_j$.

The classical fields themselves become operators:

$$\tilde{\mathbf{E}}(\mathbf{r}, t) = i\sum_{\mathbf{k},\sigma} \sqrt{\frac{\hbar\omega_{\mathrm{k}}}{2\varepsilon V}}\hat{\boldsymbol{e}}_{\mathbf{k}\sigma}\left[\tilde{a}_{\mathbf{k}\sigma}e^{i\mathbf{k}\cdot\mathbf{r}} - \tilde{a}^\dagger_{\mathbf{k}\sigma}e^{-i\mathbf{k}\cdot\mathbf{r}}\right] \tag{7.64a}$$

$$\tilde{\mathbf{B}}(\mathbf{r}, t) = i\sum_{\mathbf{k},\sigma} \sqrt{\frac{\hbar}{2\omega_{\mathrm{k}}\varepsilon V}}(\mathbf{k}\times\hat{\boldsymbol{e}}_{\mathbf{k}\sigma})\left[\tilde{a}_{\mathbf{k}\sigma}e^{i\mathbf{k}\cdot\mathbf{r}} - \tilde{a}^\dagger_{\mathbf{k}\sigma}e^{-i\mathbf{k}\cdot\mathbf{r}}\right] \tag{7.64b}$$

$$\tilde{\mathbf{A}}(\mathbf{r},t)=\sqrt{\frac{\hbar}{2\omega_{\mathrm{k}}\varepsilon V}}\sum_{\mathbf{k},\sigma}\hat{\boldsymbol{e}}_{\mathbf{k}\sigma}\left[\tilde{a}_{\mathbf{k}\sigma}e^{i\mathbf{k}\cdot\mathbf{r}}+\tilde{a}^{\dagger}_{\mathbf{k}\sigma}e^{-i\mathbf{k}\cdot\mathbf{r}}\right]. \tag{7.64c}$$

The temporal dependence of the creation and annihilation operators is given by their equations of motion for the operators in the Heisenberg picture:

$$\frac{d\tilde{a}^{\dagger}_{\mathbf{k}\sigma}}{dt}=\frac{i}{\hbar}\left[\tilde{H},\tilde{a}^{\dagger}_{\mathbf{k}\sigma}\right]=-i\omega_{\mathrm{k}}\tilde{a}^{\dagger}_{\mathbf{k}\sigma} \tag{7.65a}$$

$$\frac{d\tilde{a}_{\mathbf{k}\sigma}}{dt}=\frac{i}{\hbar}\left[\tilde{H},\tilde{a}_{\mathbf{k}\sigma}\right]=i\omega_{\mathrm{k}}\tilde{a}_{\mathbf{k}\sigma} \tag{7.65b}$$

which results in

$$\tilde{a}^{\dagger}_{\mathbf{k}\sigma}(t)=\tilde{a}^{\dagger}_{\mathbf{k}\sigma}(0)e^{-i\omega_{\mathrm{k}}t} \tag{7.66}$$

$$\tilde{a}_{\mathbf{k}\sigma}(t)=\tilde{a}_{\mathbf{k}\sigma}(0)e^{i\omega_{\mathrm{k}}t}. \tag{7.67}$$

The formal canonical quantization relies on the fundamental equal-time commutation relations between the components of the vector potential $\tilde{\mathbf{A}}(\mathbf{r},t)$ and those of its canonical conjugate momentum $\tilde{\mathbf{\Pi}}(\mathbf{r},t)=-\tilde{\mathbf{D}}(\mathbf{r},t)$:

$$\left[\tilde{\mathrm{A}}_j(\mathbf{r},t),\tilde{\Pi}_{j'}\left(\mathbf{r}',t\right)\right]=i\hbar\delta^{T}_{jj'}\left(\mathbf{r}-\mathbf{r}'\right) \tag{7.68}$$

where $\delta^{T}$is the transversal delta function[14]:

$$\delta^{T}_{jj'}\left(\mathbf{r}-\mathbf{r}'\right)=\frac{1}{(2\pi)^3}\int d\mathbf{k}\left(\delta_{jj'}-\frac{k_jk_{j'}}{k^2}\right)e^{i\mathbf{k}\cdot\left(\mathbf{r}-\mathbf{r}'\right)} \tag{7.69}$$

which acts as a delta function on transversal fields. This rigorous approach is rarely used in quantum optics but leads to the same field quantum operators (7.64a, b and c) and can be found in Cohen-Tannoudji et al. 1998.[12]

7.5 What are photons?

In 1917, Einstein published his paper on the quantum theory of radiation,[15] in which he introduced the concept of stimulated emission, discussed the momentum transfer that takes place in light emission and absorption processes by matter, and concluded that energy transfer in the amount of an energy quanta $\hbar\omega$ is always accompanied

[14]Loudon R (1973) The quantum theory of light. Oxfor Univ. Press

[15]Einstein A (1917) Phys Zeits. 18:121

by a momentum transfer $\frac{\hbar\omega}{c}$, thereby reinforcing the idea of a particle-like behavior of the energy quanta. In (7.61b) we showed that $\tilde{\mathbf{P}}$, like $\tilde{H}$, is expressible as a function of the number operators $\tilde{n}_{\mathbf{k}\lambda}$. The Fock states $|\{n_{\mathbf{k}\sigma}\}\rangle = |n_{\mathbf{k}\sigma}, n_{\mathbf{k}'\sigma'}, n_{\mathbf{k}''\sigma''}, \ldots\rangle$ are eigenstates of both $\tilde{H}$ and $\tilde{\mathbf{P}}$. Their eigenvalues mean that each of the $n_{\mathbf{k}\sigma}$ photons belonging to the $\mathbf{k}\sigma$ mode carries a momentum $\hbar\mathbf{k}$, independent of polarization σ, even though they are massless. The total momentum of the field is the sum of the momentum of the photons in each mode. The quantum field has a granular character in contrast to the continuous nature of the classical field, as if built by massless corpuscles, *photons*, that carry momentum and energy.

Photons are quantum excitations of a specific mode of the electromagnetic field. They are not particles in the classical sense, because they are in general not localized in space. The probability of detecting a photon associated with a homogenous plane wave at a specific position $\mathbf{r}$ and time t is zero. This is so because a homogenous plane wave extends over all space and does not have a beginning or and end. Notice also that if a different complete set of modes is used for the expansion of the electromagnetic field, a given eigenstate of the field in one representation may result in a superposition of states in the other. Photons and their characteristics are linked to the specific modes of the electromagnetic field that satisfy a given boundary value problem, and so is the probability to detect them in a given location. Photons in optical waveguides are more naturally associated with TE and TM modes, and photons with orbital angular momentum (OAM) with expansions in Laguerre beams.[16] A photon with a definite OAM corresponds to a superposition of plane-wave photons with $\mathbf{k}$ vectors on a conical surface. Both descriptions are accurate and are related through a unitary transformation.

The development of new micro- and nano- structured materials like photonic crystals[17] opened the possibility of achieving strong localization of light about a lattice defect.[18] The optical properties of such materials are different from those of uniform bulk dielectrics, and so are the normal modes of the electromagnetic field inside the material. A photonic crystal is a periodic lattice of macroscopic dielectric media, which can be designed to possess a photonic band gap, or range of frequencies for which light is forbidden to exist in the crystal. Photons will then be quanta of the electromagnetic field modes allowed in the crystal. The optical bandgap is the analogue of the electronic bandgap in a semiconductor that prohibits the electrons to possess energies in the bandgap. Additionally, a point-like defect in the periodicity of the photonic crystal can act as a micro-cavity and lead to a highly localized photonic state within the gap, characterized by the normal modes of the cavity rather than those of the photonic crystal.

Electromagnetic fields can also be confined in subwavelength spaces in the neighborhood of dielectric metallic interfaces that give rise to surface plasmon polaritons, which will be discussed later in this chapter.

[16] Allen L et al (1992) Phys Rev. A45:8185

[17] Yablonovitch E (1987) Phys Rev Lett. 58:2059

[18] Joannopoulos JD (1997) Nature 386:143

7.6 The Problem of Localizing Photons

In classical mechanics the state of a particle is perfectly known when its position **r** and momentum **p** are known. This is not the case in quantum mechanics in which, as described by the Heisenberg indeterminacy relations, the position and momentum of a particle cannot be measured simultaneously, and in which the state of the particle is described in the coordinate space by a wave function $\psi(\mathbf{r},t)$, a complex-valued amplitude from which the probabilities for different results of a given measurement can be derived. For one to know the position of the particle an operator position $\tilde{\mathbf{r}}$ should be applied to the wave function describing the particle

$$\tilde{\mathbf{r}}\psi(\mathbf{r},t)=\mathbf{r}\psi(\mathbf{r},t). \tag{7.70}$$

In this equation **r** is the eigenvalue of the operator $\tilde{\mathbf{r}}$, $\psi(\mathbf{r},t)$ is the probability amplitude of finding the particle at the point **r**, and its squared modulus $|\psi(\mathbf{r},t)|^2$ gives the probability.

$\psi(\mathbf{r},t)$ is a wave function for matter and can be regarded as a matter field, that can go through a second quantization process, analogous to that of the quantization of the electromagnetic field, i.e., by turning it into a quantum field operator. Second quantization allows us to know how many particles there are in a given quantum state. An example of second quantization is provided in Sect. 7.9, where we present the canonical quantization of the bulk plasmonic field.

Position operators exist in non-relativistic quantum theory for particles with any spin, and even for the relativistic theory of massive, spin ½ particles described by the Dirac equation; there is, however, no position operator for the massless, spin 1 objects described by Maxwell's equations.[19]

Over the years many proposals for a photon position operator have been made with no success. It seems that there is no physically acceptable photon position operator, consequently there is no position-space wave-function for the photon.

The nonexistence of a photon position operator still leaves the possibility that there is some other sense in which the photon may be considered as a localizable or particle-like object. It should be remembered that the number operator $\tilde{n}$ refers to the total photon number in all space. It is not expected to be accessible to direct measurement.

From a practical point of view, one could interpret the electronic signal registered by a photodetector as due to a photon that has been localized in the detector volume. More precisely, the counts registered by a detector whose surface is normal to the incident field and exposed for some finite time Δt could be interpreted as a measurement of the number of photons in a cylindrical volume whose base is the sensitive surface of the detector and whose height is $c\Delta t$. The integral of the intensity I(r,t) over such volume can therefore be interpreted as a configuration space number operator, at least in an approximate sense, provided the linear

[19] Wigner TD, Newton EP (1949) Rev Mod Phys. 21:400

dimensions of the volume are all large compared with the wavelengths of contributing modes (see for example Mandel and Wolf[20] for a detailed discussion).

This is of course a very simplified point of view. Briefly, a fundamental point is that a position operator cannot exist in a relativistic quantum theory. Position operators exist in non-relativistic quantum theory for particles with any spin, and even for relativistic massive spin ½ particles described by the Dirac equation, but there is no position operator for the massless, spin 1 photons. Even for massive particles the localization is not perfect, because it is not relativistically invariant. Two observers in relative motion would not quite agree as to the localization region of a relativistic particle.

We need not to enter in this matter and refer, among others, to review papers by Acharya and Sudarshan,[21] Bialynicki-Birula,[22] Scully and Zubairy,[23] and O. Keller.[24] We assume the pragmatic position that a single-photon was there if a detector, that can determine the number of incident photons with 100% accuracy, clicks once.

7.7 Expansion of the Field in Orthonormal Modes

The increased ability of fabricating nanostructures with periodical heterostructures or with embedded nanoparticles has brought out the need for performing quantization of the electromagnetic field in small volumes using the modal structure of the field that satisfies the specific boundary conditions of the problem rather than plane waves. The simplest case, that of a one-dimensional Fabry-Perot cavity was described in the first section. When dealing with surface plasmons, the classical model consists of two homogeneous media characterized by local, frequency-dependent, dielectric functions $\varepsilon_1$and ε_2 and separated by an interface of arbitrary geometry. Maxwell's equations are then solved with the appropriate boundary conditions as it was done in Chap. V for a planar interface. The same analytical procedure can be used for localized structures with simple geometrical shapes, such as spheres and cylinders,[25] but the electromagnetic modes of arbitrary geometric shapes usually need to be determined numerically.

The quantization of the electromagnetic field confined in a cavity or to a region about a metal-dielectric interface requires in general the determination of the classical modes that satisfy Maxwell's equations and their normalization. We will express all classical fields as complex functions of position and time, that can be derived from the vector potential $\mathbf{A}_\mu(\mathbf{r}, t)$ of the classical mode μ. The label μ represents all

[20]Mandel L, Wolf E (1995) Optical coherence and quantum optics. Cambridge Univ. Press

[21]Acharia R, Sudarshan ECG (1960) J Math Phys. 1:532

[22]Bialynicki-Birula IV (1996) Prog Opt. 36:245

[23]Scully MO, Zubairy MS (1997) Quantum optics. Cambridge Univ. Press

[24]Keller O (2007) Progr Opt. 50:51

[25]Pitarke JM et al (2007) Rep Prog Phys. 70:1

physical quantities required to specify the mode. The complex conjugate $\mathbf{A}_\mu^*(\mathbf{r},t)$ should always be added to the complex field because Maxwell equations are real.

To orthonormalize the complex field $\mathbf{A}_\mu(\mathbf{r},t)$ we use the scalar product in the function space of the modes as defined in[26]:

$$\left(\mathbf{A}_\mu, \mathbf{A}_{\mu'}\right) \equiv \frac{1}{i\hbar}\int d\mathbf{r}\left(\mathbf{A}_\mu^* \cdot \mathbf{D}_{\mu'} - \mathbf{A}_{\mu'} \cdot \mathbf{D}_\mu^*\right) \tag{7.71a}$$

$$\mathbf{D}_\mu = -\varepsilon_0 \varepsilon \frac{\partial \mathbf{A}_\mu}{\partial t}, \tag{7.71b}$$

where the factor $1/i\hbar$ was introduced for convenience. We recall that $-\mathbf{D}$ is the canonical conjugate momentum of the vector potential. The scalar product defined in (7.71a) shares some of the algebraic features of commutators:

$$\left(\mathbf{A}_\mu^*, \mathbf{A}_{\mu'}^*\right) = -\left(\mathbf{A}_\mu, \mathbf{A}_{\mu'}\right)^*. \tag{7.72}$$

It is not positive definite and its norm $\left(\mathbf{A}_\mu, \mathbf{A}_\mu\right)$ is zero, but it is a conserved quantity, i.e.,

$$\frac{d}{dt}\left(\mathbf{A}_\mu, \mathbf{A}_{\mu'}\right) = 0. \tag{7.73}$$

This extra symmetry of complex fields, not owned by real fields, is related to the fact that complex fields obey real equations. If the complex mode fields are multiplied by a global phase factor, the new node obeys the same wave equations.

Complex field modes that are orthonormal with respect to the scalar product (7.71a) are called *normal* modes. They satisfy

$$\left(\mathbf{A}_\mu, \mathbf{A}_{\mu'}\right) = \delta_{\mu\mu'}, \qquad \left(\mathbf{A}_\mu, \mathbf{A}_{\mu'}^*\right) = 0. \tag{7.74}$$

The vector field operator $\tilde{\mathbf{A}}(\mathbf{r},t)$ can be expanded in terms of the orthonormal modes as

$$\tilde{\mathbf{A}}(\mathbf{r},t) = \sum_\mu \left[\mathbf{A}_\mu(\mathbf{r},t)\tilde{a}_\mu + \mathbf{A}_\mu^*(\mathbf{r},t)\tilde{a}_\mu^\dagger\right]. \tag{7.75}$$

The mode operators $\tilde{a}_\mu$ and $\tilde{a}_\mu^\dagger$ are the projections of $\tilde{\mathbf{A}}(\mathbf{r},t)$ into the normal modes:

$$\tilde{a}_\mu = \left(\mathbf{A}_\mu, \tilde{\mathbf{A}}\right) \tag{7.76a}$$

$$\tilde{a}_\mu^\dagger = -\left(\mathbf{A}_\mu^*, \tilde{\mathbf{A}}\right). \tag{7.76b}$$

[26] Leonhardt U (2010) Essential quantum optics. Cambridge Univ. Press

Using the orthonormality conditions (7.74) and the fundamental commutation relation (7.68), we can obtain the bosonic commutation relations,

$$\left[\tilde{a}_\mu, \tilde{a}^\dagger_{\mu'}\right] = \delta_{\mu\mu'} \tag{7.77a}$$

$$\left[\tilde{a}^\dagger_{\mu'}, \tilde{a}^\dagger_\mu\right] = \left[\tilde{a}_{\mu'}, \tilde{a}_\mu\right] = 0 \tag{7.77b}$$

which characterize the annihilation and creation operators $\hat{a}_\mu$ and $\hat{a}^\dagger_\mu$.

The field expansion (7.75) is not restricted to monochromatic fields. It can be used for light pulses. For the special case of monochromatic fields,

$$\frac{\partial \mathbf{A}_\mu(\mathbf{r},t)}{\partial t} = -i\omega_\mu \mathbf{A}_\mu(\mathbf{r},t) \tag{7.78a}$$

$$\mathbf{D}_\mu = -\varepsilon_0 \varepsilon \frac{\partial \mathbf{A}_\mu}{\partial t} = i\omega_\mu \varepsilon_0 \varepsilon \mathbf{A}_\mu(\mathbf{r},t), \tag{7.78b}$$

and the scalar product (7.71a and b) becomes

$$(\mathbf{A}_\mu, \mathbf{A}_{\mu'}) = \frac{2\varepsilon_0 \omega_\mu}{\hbar} \int d\mathbf{r}\, \varepsilon(\mathbf{r}) \mathbf{A}^*_\mu \cdot \mathbf{A}_{\mu'}, \tag{7.79}$$

which, except for the inclusion of the dielectric function in the integral, resembles the normalization integral for a wavefunction in quantum mechanics. This expression allows the proper normalization of mode fields in inhomogeneous media, and the normalization of evanescent waves, surface plasmon polaritons and localized SPP. We assume that the quantum field operators satisfy the same linear differential equations and boundary conditions that are satisfied by the classical fields. Therefore, the electric field operator for monochromatic fields can be obtained from (7.75) by differentiation with respect to time:

$$\tilde{\mathbf{E}}(\mathbf{r},t) = i\omega_\mu \sum_\mu \left[\mathbf{A}_\mu(\mathbf{r},t)\tilde{a}_\mu - \mathbf{A}^*_\mu(\mathbf{r},t)\tilde{a}^\dagger_\mu\right]. \tag{7.80}$$

Monochromatic modes describe stationary processes that conserve energy. The electromagnetic energy

$$H = \frac{1}{2}\int d\mathbf{r}(\mathbf{E}\cdot\mathbf{D} + \mathbf{B}\cdot\mathbf{H}) \tag{7.81}$$

should then be time independent. For localized fields we can use the following identity,

$$\mathbf{A}\cdot\frac{\partial \mathbf{D}}{\partial t} = \mathbf{A}\cdot(\nabla\times\mathbf{H}) = -\nabla\cdot(\mathbf{A}\times\mathbf{H}) + \mathbf{B}\cdot\mathbf{H}, \tag{7.82}$$

to replace the second term in the integral (7.81). According to Gauss's theorem, the integral of the term $\nabla \cdot (\mathbf{A} \times \mathbf{H})$ becomes a surface integral of $\mathbf{A} \times \mathbf{H}$ at infinity where the fields vanish, and the Hamiltonian (7.81) becomes

$$H = \frac{1}{2} \int d\mathbf{r} \left(\mathbf{E} \cdot \mathbf{D} + \mathbf{A} \cdot \frac{\partial \mathbf{D}}{\partial t} \right). \tag{7.83}$$

Assuming the Hamiltonian operator has the same structure as does the classical energy of the electromagnetic field (7.81), we can write

$$\tilde{H} = \frac{1}{2} \int d\mathbf{r} \left(\tilde{\mathbf{E}} \cdot \tilde{\mathbf{D}} + \tilde{\mathbf{A}} \cdot \frac{\partial \tilde{\mathbf{D}}}{\partial t} \right). \tag{7.84}$$

Using (7.75), (7.78b), (7.79), (7.80), and the orthogonality relation (7.74) we find

$$\tilde{H} = \sum_{\mu} \frac{\hbar \omega_{\mu}}{2} \left(\tilde{a}_{\mu} \tilde{a}_{\mu}^{\dagger} + \tilde{a}_{\mu}^{\dagger} \tilde{a}_{\mu} \right). \tag{7.85}$$

Finally, using the commutation relation (7.77a), we obtain

$$\tilde{H} = \sum_{\mu} \hbar \omega_{\mu} \left(\tilde{a}_{\mu} \tilde{a}_{\mu}^{\dagger} + \frac{1}{2} \right). \tag{7.86}$$

The total energy corresponds to the sum of the energies of independent harmonic oscillators, each associated with a mode μ.

7.8 Quantization of Evanescent Waves

Propagation in free space or bulk uniform dielectrics does not account for light traversing lenses or other optical elements. Even the analysis of refraction through a flat interface would require the consideration of the interaction of the free-space quantized field with the material as a coupled system. The canonical quantization based on the expansion of the field in homogenous plane waves is not valid if, for example, sources, scatterers, or apertures are present, and it admits no evanescent waves components arising from total internal reflection.

Instead of introducing the material medium and its interaction with the electromagnetic field explicitly, Carniglia and Mandel (1971)[1] demonstrated the quantization of the electromagnetic field when the half-space to the left of the plane at $z = 0$ is filled with a homogenous dielectric with refractive index n and the half-space to the right is empty (Fig. 7.3). The refractive index is assumed to be real, which is a good approximation far from resonances. They treated the transmitted homogenous or evanescent wave together with the incident and reflected waves as one single mode, known as a *triplet mode*.

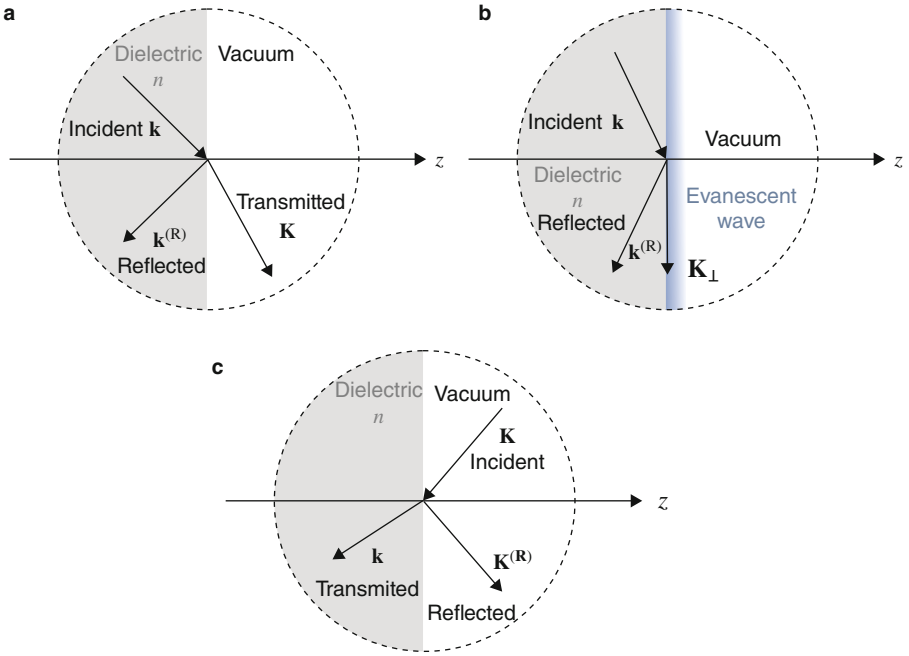


Fig. 7.3 (**a**) A wave incident on the dielectric-vacuum interface from the left is partially reflected and partially transmitted. The three waves—incident, reflected and transmitted—constitute a triplet mode. (**b**) Triplet mode for incidence from the left giving rise to an evanescent wave. (**c**) Triplet mode for incidence from the right

A triplet mode with incident wave from the left is labeled by the wave vector $\mathbf{k} = (\mathrm{k}_x, \mathrm{k}_y, \mathrm{k}_z)$, and its TE or TM polarization inside the dielectric. Triplet modes with incidence from the right (Fig. 7.3c) are labeled by the incident wave vector $\mathbf{K} = (\mathrm{K}_x, \mathrm{K}_y, \mathrm{K}_z)$ in the vacuum. The magnitudes and components of the wave vector **K** are related to those of **k** by:

$$\mathrm{K} = \frac{\mathrm{k}}{n} = \frac{\omega}{c} \tag{7.87a}$$

$$\mathrm{K}_x = \mathrm{k}_x \tag{7.87b}$$

$$\mathrm{K}_y = \mathrm{k}_y. \tag{7.87c}$$

For left incidence:

$$\mathrm{K}_z = \pm\sqrt{\mathrm{K}^2 - \mathrm{K}_x^2 - \mathrm{K}_y^2} \tag{7.88a}$$

$$\mathrm{k}_z = \pm\sqrt{n^2\mathrm{K}^2 - \mathrm{k}_x^2 - \mathrm{k}_y^2} \tag{7.88b}$$

k is always a real vector but for left incident light **K** will be complex when $K^2 < K_x^2 + K_y^2$. Its z-component K_z becomes purely imaginary leading to an evanescent wave, whose amplitude decays exponentially with distance from the interface, while propagating parallel to the interface with the wave vector $\mathbf{K}_{\parallel} = (K_x, K_y)$. For incidence from the right (Fig. 7.3c), there is always a transmitted homogeneous wave.

The reflected and transmitted fields are related to the incident field by means of the Fresnel relations (see Sect. 2.2.3) for the interface, and they exist only in the dielectric and the vacuum respectively, i.e., the incident and reflected fields are set to zero outside the dielectric ($z \geq 0$), and the homogenous transmitted (or evanescent) field is set to zero inside the dielectric ($z < 0$). The sum of the three fields satisfies the Helmholtz equation:

$$\nabla^2 \mathbf{E} + K^2 n(r) \mathbf{E} = 0, \quad n(r) = \begin{cases} n > 1, & z < 0 \\ 1, & z \geq 0 \end{cases}. \tag{7.89}$$

7.8.1 Triplet Mode Functions

To quantize evanescent waves, after Carniglia and Mandel and T. Inoue and H. Hori,[27] we need to consider both the L-triplet mode functions (Fig. 7.4a) for light incident from the left, and the R-triplet mode functions for light incident from the right (Fig. 7.4b). There are two different sets of L- and R-triplet modes for TE and TM polarization, which will be labeled with the sub index $\mu = 1, 2$, respectively. The electric field for both, L and R triplets, consist of a composition of incident, $\mathbf{E}^{(I)}_{L(R)\mu}$, reflected, $\mathbf{E}^{(R)}_{L(R)\mu}$ and transmitted fields $\mathbf{E}^{(T)}_{L(R)\mu}$:

$$\mathbf{E}_{L\mu}(\mathbf{k}, \mathbf{r}) = \mathbf{E}^{(I)}_{L\mu}(\mathbf{k}, \mathbf{r}) + \mathbf{E}^{(R)}_{L\mu}\left(\mathbf{k}^{(R)}, \mathbf{r}\right) + \mathbf{E}^{(T)}_{L\mu}(\mathbf{K}, \mathbf{r}). \tag{7.90}$$

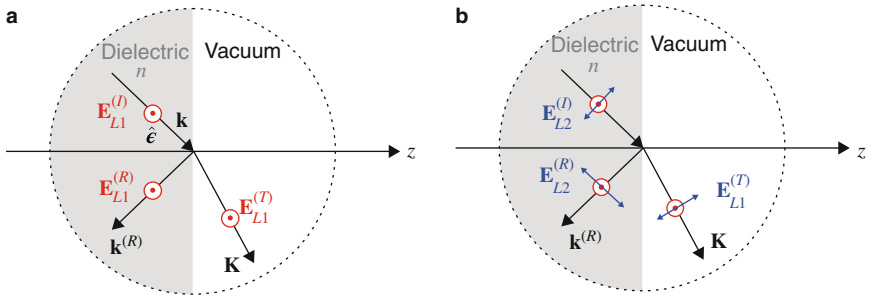


Fig. 7.4 (**a**) Left triplet TE mode. The Electric field is parallel to the interface, pointing along the unit vector $\hat{\epsilon}$. (**b**) Left triplet TM mode. The magnetic field is parallel to the interface and points along $\hat{\epsilon}$. Note that the electric fields for the TE and TM modes are orthogonal

[27] Hinoue T, Hori H (2001) Phys Rev. A63:063805

$$\mathbf{E}_{R\mu}(\mathbf{k},\mathbf{r}) = \mathbf{E}^{(I)}_{R\mu}(\mathbf{K},\mathbf{r}) + \mathbf{E}^{(R)}_{R\mu}\left(\mathbf{K}^{(\boldsymbol{R})},\mathbf{r}\right) + \mathbf{E}^{(T)}_{R\mu}(\mathbf{k},\mathbf{r}) \tag{7.91}$$

The field and wavevector directions for TE and TM L-triplet modes are shown in Fig. 7.4a,b.

The fields for the left L-triplet mode are defined in Table 7.1, where $\mathrm{E}_0 = \sqrt{4\pi^3\varepsilon_0/\hbar\omega}$ is a normalization constant that has been chosen such that the energy density contained in a triplet mode, when integrated over the whole space, corresponds to that of a single photon of frequency $\hbar\omega$. This condition is consistent with the orthonormality condition (7.74) and the definition of the scalar product (7.79). The light source is on the left, and $\mathbf{k}$ and $\mathbf{k}^{(\boldsymbol{R})}$ are real vectors. From (7.88a), the component K_z of the wave vector $\mathbf{K}$ of the transmitted wave is a real number for $0 \le \mathrm{k}_x^2 + \mathrm{k}_y^2 < \mathrm{K}^2$, but a complex number for $\mathrm{K}^2 \le \mathrm{k}_x^2 + \mathrm{k}_y^2 < n^2\mathrm{K}^2$ (see Table 7.1).

The electric field for the right R-triplet mode functions (Fig. 7.3c) are given in Table 7.2.

Since for R-triplet modes, the light is incident from the right, $\mathbf{E}^{(I)}_{R1}(\mathbf{K},\mathbf{r})$ corresponds to a homogenous wave with real $\mathbf{K}$ vector. The corresponding wave-vector component k_z for the transmitted wave in the left side is always real since $0 \le \mathrm{k}_x^2 + \mathrm{k}_y^2 = \mathrm{K}_x^2 + \mathrm{K}_y^2 < \mathrm{K}^2$. Therefore all the wave-vectors in the right triplet mode are real.

Carniglia and Mandel (1971) demonstrated the orthonormality of the L- and R-triplet mode functions (7.90) and (7.91). In agreement with (7.79) the orthonormality relations can be written as

$$\frac{2\varepsilon_0}{\hbar\omega}\int \mathbf{E}^{*}_{L\mu}(\mathbf{k},\mathbf{r})\cdot\mathbf{E}_{L\mu'}(\mathbf{k}',\mathbf{r})\;\; n^2(\mathbf{r})d\mathbf{r} = \delta_{\mu\mu'}\delta(\mathbf{k}-\mathbf{k}'). \tag{7.92}$$

Table 7.1 Electric field for the L-triplet modes

		$z<0$	$z\ge 0$
TE	$\mathbf{E}^{(I)}_{L1}(\mathbf{k},\mathbf{r})$	$\frac{\mathrm{E}_0}{\sqrt{2n}}\hat{\boldsymbol{\epsilon}}\; e^{i\mathbf{k}\cdot\mathbf{r}}$	0
	$\mathbf{E}^{(R)}_{L1}\left(\mathbf{k}^{(\boldsymbol{R})},\mathbf{r}\right)$	$\frac{\mathrm{E}_0}{\sqrt{2n}}\hat{\boldsymbol{\epsilon}}\,\frac{\mathrm{k}_z+\mathrm{K}_z}{\mathrm{k}_z-\mathrm{K}_z}e^{i\mathbf{k}^{(\boldsymbol{R})}\cdot\mathbf{r}}$	0
	$\mathbf{E}^{(T)}_{L1}(\mathbf{K},\mathbf{r})$	0	$\frac{\mathrm{E}_0}{\sqrt{2n}}\hat{\boldsymbol{\epsilon}}\frac{2\mathrm{k}_z}{\mathrm{k}_z-\mathrm{K}_z}e^{i\mathbf{K}\cdot\mathbf{r}}$
TM	$\mathbf{E}^{(I)}_{L2}(\mathbf{k},\mathbf{r})$	$-\frac{\mathrm{E}_0}{\sqrt{2n}}(\hat{\mathbf{k}}\times\hat{\boldsymbol{\epsilon}})e^{i\mathbf{k}\cdot\mathbf{r}}$	0
	$\mathbf{E}^{(R)}_{L2}\left(\mathbf{k}^{(\boldsymbol{R})},\mathbf{r}\right)$	$-\frac{\mathrm{E}_0}{\sqrt{2n}}(\hat{\mathbf{k}}^{(\boldsymbol{R})}\times\hat{\boldsymbol{\epsilon}})\frac{\mathrm{k}_z+n^2\mathrm{K}_z}{\mathrm{k}_z-n^2\mathrm{K}_z}e^{i\mathbf{k}^{(\boldsymbol{R})}\cdot\mathbf{r}}$	0
	$\mathbf{E}^{(T)}_{L2}(\mathbf{K},\mathbf{r})$	0	$-\frac{\mathrm{E}_0}{\sqrt{2n}}(\hat{\mathbf{K}}\times\hat{\boldsymbol{\epsilon}})\frac{2n\mathrm{k}_z}{\mathrm{k}_z-n^2\mathrm{K}_z}e^{i\mathbf{K}\cdot\mathbf{r}}$

Table 7.2 Electric field for the R-triplet modes

		$z \geq 0$	$z < 0$
TE	$\mathbf{E}_{R1}^{(I)}(\mathbf{K}, \mathbf{r})$	$\frac{\mathrm{E}_0}{\sqrt{2}}\hat{\boldsymbol{\epsilon}}\, e^{i\mathbf{K}\cdot\mathbf{r}}$	0
	$\mathbf{E}_{R1}^{(R)}(\mathbf{K}^{(R)}, \mathbf{r})$	$\frac{\mathrm{E}_0}{\sqrt{2n}}\hat{\boldsymbol{\epsilon}}\frac{\mathrm{K}_z + \mathrm{k}_z}{\mathrm{K}_z - \mathrm{k}_z} e^{i\mathbf{K}^{(R)}\cdot\mathbf{r}}$	0
	$\mathbf{E}_{R1}^{(T)}(\mathbf{k}, \mathbf{r})$	0	$\frac{\mathrm{E}_0}{\sqrt{2}}\hat{\boldsymbol{\epsilon}}\frac{2\mathrm{K}_z}{\mathrm{K}_z - \mathrm{k}_z} e^{i\mathbf{k}\cdot\mathbf{r}}$
TM	$\mathbf{E}_{R2}^{(I)}(\mathbf{K}, \mathbf{r})$	$-\frac{\mathrm{E}_0}{\sqrt{2}}(\hat{\mathbf{K}} \times \hat{\boldsymbol{\epsilon}}) e^{i\mathbf{K}\cdot\mathbf{r}}$	0
	$\mathbf{E}_{R2}^{(R)}(\mathbf{K}^{(R)}, \mathbf{r})$	$-\frac{\mathrm{E}_0}{\sqrt{2}}(\hat{\mathbf{K}}^{(R)} \times \hat{\boldsymbol{\epsilon}})\frac{n^2\mathrm{K}_z + \mathrm{k}_z}{n^2\mathrm{K}_z - \mathrm{k}_z} e^{i\mathbf{K}^{(R)}\cdot\mathbf{r}}$	0
	$\mathbf{E}_{R2}^{(T)}(\mathbf{k}, \mathbf{r})$	0	$-\frac{\mathrm{E}_0}{\sqrt{2n}}(\hat{\mathbf{k}} \times \hat{\boldsymbol{\epsilon}})\frac{2n\mathrm{K}_z}{n^2\mathrm{K}_z - \mathrm{k}_z} e^{i\mathbf{k}\cdot\mathbf{r}}$

$$\frac{2\varepsilon_0}{\hbar\omega}\int \mathbf{E}_{R\mu}^{*}(\mathbf{K}, \mathbf{r}) \cdot \mathbf{E}_{R\mu'}\left(\mathbf{K}', \mathbf{r}\right) n^2(\mathbf{r}) d\mathbf{r} = \delta_{\mu\mu'}\delta\left(\mathbf{K} - \mathbf{K}'\right) \tag{7.93}$$

$$\int \mathbf{E}_{L\mu}^{*}(\mathbf{k}, \mathbf{r}) \cdot \mathbf{E}_{R\mu'}(\mathbf{K}', \mathbf{r})\, n^2(\mathbf{r}) d\mathbf{r} = 0. \tag{7.94}$$

The completeness of the triplet modes was demonstrated by I. Białynicki-Birula and J. B. Brojan.[28] The modes thus constitute an appropriate basis set for expansion of any arbitrary electromagnetic field as

$$\mathbf{E}(\mathbf{r}, t) = \sum_{\mu=1}^{2} \mathbf{E}_{\mu}^{+}(\mathbf{r}, t) e^{-i\omega t} + c.c. \tag{7.95}$$

The amplitude of the modal functions is given by

$$\mathbf{E}_{\mu}^{+}(\mathbf{r}, t) = \int d\mathbf{k}_{\parallel} \int_0^{\infty} d\mathrm{k}_z \alpha_{L\mathbf{k}\mu} \mathbf{E}_{L\mu}(\mathbf{k}, \mathbf{r}) + \int d\mathbf{K}_{\parallel} \int_{-\infty}^{0} d\mathrm{K}_z \alpha_{R\mathbf{K}\mu} \mathbf{E}_{R\mu}(\mathbf{K}, \mathbf{r}). \tag{7.96}$$

The integrals in (7.96) are defined as follows:

$$\int d\mathbf{k}_{\parallel} = \int_{-\infty}^{\infty} dk_x \int_{-\infty}^{\infty} dk_y, \tag{7.97a}$$

$$\int d\mathbf{K}_{\parallel} = \int_{-\infty}^{\infty} dK_x \int_{-\infty}^{\infty} dK_y \tag{7.97b}$$

[28] Bialynicki-Birula I, Brojan JB (1972) Phys Rev. D5:485

The magnetic fields associated with each triplet mode follow directly from applying Maxwell (1.1c) to the electric fields in Tables 7.1 and 7.2. Thus

$$\mathbf{B}(\mathbf{r},t)=\sum_{\mu=1}^{2}\mathbf{B}_{\mu}^{+}(\mathbf{r},t)e^{-i\omega t}+c.c. \tag{7.98a}$$

$$\mathbf{B}_{\mu}^{+}(\mathbf{r},t)=\frac{1}{\omega}\left[\int d\mathbf{k}_{\parallel}\int_{0}^{\infty}d\mathrm{k}_{z}\alpha_{L\mathbf{k}\mu}\mathbf{k}\times\mathbf{E}_{L\mu}(\mathbf{k},\mathbf{r})+\int d\mathbf{K}_{\parallel}\int_{-\infty}^{0}d\mathrm{K}_{z}\alpha_{R\mathbf{K}\mu}\mathbf{K}\times\mathbf{E}_{R\mu}(\mathbf{K},\mathbf{r})\right]. \tag{7.98b}$$

The total energy of the electromagnetic field is given by

$$H=\frac{1}{2}\int\left[\varepsilon_{0}n_{0}^{2}(\mathbf{r})\mathbf{E}(\mathbf{r},t)\cdot\mathbf{E}(\mathbf{r},t)+\frac{\mathbf{B}^{2}(\mathbf{r},t)}{\mu_{0}}\right]d\mathbf{r}\,. \tag{7.99}$$

Using the orthogonality relations (7.92, 7.93 and 7.94) we obtain a simple expression for the Hamiltonian of the electromagnetic field:

$$H=\frac{\hbar\omega}{2}\sum_{\mu=1}^{2}\int d\mathbf{k}_{\parallel}\int_{0}^{\infty}d\mathrm{k}_{z}\left(\alpha_{L\mathbf{k}\mu}^{*}\alpha_{L\mathbf{k}\mu}+c.c.\right)+\int d\mathbf{K}_{\parallel}\int_{-\infty}^{0}d\mathrm{K}_{z}\left(\alpha_{R\mathbf{K}\mu}^{*}\alpha_{R\mathbf{K}\mu}+c.c.\right). \tag{7.100}$$

7.8.2 *Field Quantization Based on Triplet Mode Functions*

For planar boundaries, the complete set of triplet modes can be used as a "source-free" representation of the electromagnetic field in the combined dielectric and vacuum half spaces. The quantization of the electromagnetic energy is achieved by turning the expansion coefficients $\alpha_{L\mathbf{k}\mu}$ and $\alpha_{R\mathbf{K}\mu}$ in (7.100) into photon annihilation-operators $\tilde{a}_{L\mathbf{k}\mu}$ and $\hat{a}_{R\mathbf{K}\mu}$ for the L- and R-triplet modes. The complex conjugate amplitudes are replaced by the Hermitian conjugate operators $\tilde{a}_{L\mathbf{k}\mu}^{\dagger}$ and $\tilde{a}_{R\mathbf{K}\mu}^{\dagger}$, or photon creation operators.

The quantum-mechanical operators satisfy the following commutation relations:

$$\left[\tilde{a}_{L\mathbf{k}\mu},\tilde{a}_{L\mathbf{k}'\mu'}^{\dagger}\right]=\delta_{\mu\mu'}\delta\left(\mathbf{k}-\mathbf{k}'\right) \tag{7.101a}$$

$$\left[\tilde{a}_{R\mathbf{K}\mu},\tilde{a}_{R\mathbf{K}'\mu'}^{\dagger}\right]=\delta_{\mu\mu'}\delta\left(\mathbf{K}-\mathbf{K}'\right) \tag{7.101b}$$

$$\left[\tilde{a}_{R\mathbf{K}\mu},\tilde{a}_{L\mathbf{k}'\mu'}\right]=0, \tag{7.101c}$$

$$\left[\tilde{a}_{R\mathbf{K}\mu}, \tilde{a}^{\dagger}_{L\mathbf{k}'\mu'}\right] = 0. \tag{7.101d}$$

The Hamiltonian operator becomes:

$$\hat{H} = \hbar\omega \sum_{\mu=1}^{2} \left\{ \int d\mathbf{k}_{\parallel} \int_{0}^{\infty} d\mathrm{k}_z \left[\tilde{a}^{\dagger}_{L\mathbf{k}\mu} \tilde{a}_{L\mathbf{k}\mu} + \frac{1}{2} \right] + \int d\mathbf{K}_{\parallel} \int_{-\infty}^{0} d\mathrm{K}_z \left[\tilde{a}^{\dagger}_{R\mathbf{K}\mu} \tilde{a}_{R\mathbf{K}\mu} + \frac{1}{2} \right] \right\}, \tag{7.102}$$

which can be written in terms of the photon number operators for the L- and R-triplet modes,

$$\tilde{N}_{L\mathbf{k}\mu} = \tilde{a}^{\dagger}_{L\mathbf{k}\mu} \tilde{a}_{L\mathbf{k}\mu}, \tag{7.103a}$$

$$\tilde{N}_{R\mathbf{K}\mu} = \tilde{a}^{\dagger}_{R\mathbf{K}\mu} \tilde{a}_{R\mathbf{K}\mu}, \tag{7.103b}$$

with the result

$$\hat{H} = \hbar\omega \sum_{\mu=1}^{2} \left\{ \int d\mathbf{k}_{\parallel} \int_{0}^{\infty} d\mathrm{k}_z \left[\tilde{N}_{L\mathbf{k}\mu} + \frac{1}{2} \right] + \int d\mathbf{K}_{\parallel} \int_{-\infty}^{0} d\mathrm{K}_z \left[\tilde{N}_{R\mathbf{K}\mu} + \frac{1}{2} \right] \right\}. \tag{7.104}$$

The number eigenstates of the L- or R-triplet modes, can be generated by applying the creation operators $\tilde{a}^{\dagger}_{L\mathbf{k}\mu}$ or $\tilde{a}^{\dagger}_{R\mathbf{K}\mu}$ to a vacuum state $|0\rangle$, with the normalization $\langle 0|0\rangle = 1$. A quantum state with n photons in the L-triplet mode $\mathbf{E}_{L\mu}(\mathbf{k}, \mathbf{r})$ and m photons in the R-triplet mode $\mathbf{E}_{R\mu'}(\mathbf{K}, \mathbf{r})$ will be obtained from the vacuum state by

$$\frac{\left(\tilde{a}^{\dagger}_{L\mathbf{k}\mu}\right)^{n} \left(\tilde{a}^{\dagger}_{R\mathbf{K}\mu'}\right)^{m}}{\sqrt{n!m!}} |0\rangle = \left| n_{L\mathbf{k}\mu}, m_{R\mathbf{K}\mu'} \right\rangle. \tag{7.105}$$

The total energy of the electromagnetic field is expressible as the sum of contributions from independent harmonic oscillators, one for each triplet mode. The quantization of the field is then essentially the same as that for the free field in vacuum, except for the fact that the fundamental modes are different, and that each mode, although labelled by one wave vector and one polarization index, always corresponds to three waves coupled via the Fresnel formulas, one of which (for $\mathrm{k}_z > 0$) may be an evanescent wave. Since the evanescent waves constitute only a component of the triplet mode and extend only over a half-space for a limited range of triplet modes, there are no completely independent evanescent photons. Any triplet mode that includes an evanescent wave includes also homogenous plane waves in the left half-space. Stated differently, there exists no evanescent photon by itself. A photon in the L-triplet mode can be detected either in the reflected or in the evanescent wave. The probabilities of detection in one or the other are already immersed in the structure of the triplet mode.

Triplet modes were demonstrated to lead to the same results as the semiclassical methods for the description of excitation of molecules by evanescent light and the emission of evanescent light in the fluorescence of excited molecules.[29] With the advent of atom optics and optical near-field microscopy and spectroscopy, the understanding of photoemission processes of atoms and molecules in the proximity of material surfaces became fundamental. Evanescent waves have been used to manipulate cold atoms and confine them in atomic resonators, and near-field techniques involve radiating systems at subwavelength distances from matter, including those of quantum dots and wires on a substrate. The use of triplet modes and similar expansions (Hinoue, Hori, 2001)[26] has contributed to the understanding of such phenomena and highlighted the fact that the interaction takes place with the complete mode and not just with its evanescent wave component.

7.9 Plasmons in Bulk Metals

The valence electrons in a metallic solid can be thought as of being detached from the ions, thereby constituting an electron plasma, or a free electron gas. Plasma oscillations of the free electron gas density are analogous to a sound wave, which is a density wave in a gas consisting of molecules (Fig. 7.5). They are collective longitudinal density waves of an electron gas with respect to the positive ions. Plasma oscillations are also quantized, and their energy quanta are called *plasmons*, or more specifically for a bulk material, volume or *bulk plasmons* (to distinguish them from *surface plasmons*, which are discussed in the next section).

In the ground state of the crystal the ions are "frozen" in their equilibrium positions (except for quantum fluctuations) and the electron gas is in the lowest-energy state, with the electrons filling all states up to the Fermi level. When electrons

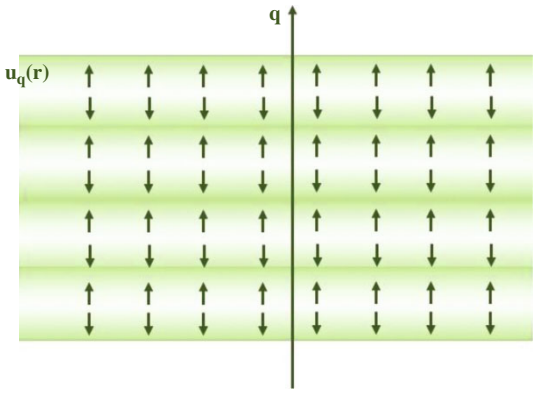


Fig. 7.5 A plasma oscillation consists of alternating regions of enhanced and rarefied plasma charge density. Plasma oscillations are longitudinal. The displacements $\mathbf{u_q}(\mathbf{r})$ of the charges from their equilibrium positions (shown as arrows) oscillate with the plasma frequency along the direction of propagation, i.e., along the wave vector $\mathbf{q}$

[29] Carniglia CK et al (1972) J Opt Soc America 62:479

acquire additional energy, usually through an external perturbation, Coulomb forces act as restoring forces and originate plasma oscillations. Displacements $\mathbf{u}(\mathbf{r})$ of the electron gas at position $\mathbf{r}$, can be expanded in its normal modes through a Fourier decomposition:

$$\mathbf{u}(\mathbf{r}) = \sum_{\mathbf{q}} \mathbf{u}_{\mathbf{q}}(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}}, \tag{7.106}$$

where $\mathbf{u}_{\mathbf{q}}(\mathbf{r})$ is the canonical coordinate (displacements from equilibrium position at $\mathbf{r}$) for the $\mathbf{q}$-mode. Since plasma oscillations are longitudinal, its normal modes are fully characterized by its wave vector $\mathbf{q}$. The canonical conjugate momenta are

$$\boldsymbol{\pi}_{\mathbf{q}} = N_0 m \dot{\mathbf{u}}_{\mathbf{q}}^*, \tag{7.107}$$

where N_0 is the uniform free electron density, and m is the effective electron mass. If deviations from the equilibrium are small (small displacements), restoring forces can be considered to be harmonic and the Hamiltonian results in a sum of independent Hamiltonians of single harmonic oscillators, each one corresponding to a normal mode:

$$H_{plasma} = \sum_{\mathbf{q}} \frac{1}{2N_0 m} \left(\boldsymbol{\pi}_{\mathbf{q}}^* \cdot \boldsymbol{\pi}_{\mathbf{q}} + \omega_P^2 \, \mathbf{u}_{\mathbf{q}}^* \cdot \mathbf{u}_{\mathbf{q}} \right), \tag{7.108}$$

where ω_P is the plasma frequency of the metal. Canonical quantization follows straightforwardly, the canonical coordinates and momenta being treated as operators and bosonic commutation relations being introduced between its components:

$$\left[\tilde{u}_{\mathbf{q},i}, \tilde{\pi}_{\mathbf{q}',j} \right] = i\hbar \, \delta_{\mathbf{q},\mathbf{q}'} \delta_{i,j}. \tag{7.109}$$

Plasmon annihilation and creation operators are introduced via the relations:

$$\tilde{\mathbf{u}}_{\mathbf{q}} = \left(\frac{\hbar}{2N_0 m} \right)^{\frac{1}{2}} \left(\tilde{\mathbf{b}}_{\mathbf{q}} + \tilde{\mathbf{b}}_{\mathbf{q}}^{\dagger} \right) \tag{7.110}$$

$$\hat{\boldsymbol{\pi}}_{\mathbf{q}} = i \left(\frac{\hbar N_0 m \omega_P}{2} \right)^{\frac{1}{2}} \left(\tilde{\mathbf{b}}_{\mathbf{q}} - \tilde{\mathbf{b}}_{\mathbf{q}}^{\dagger} \right) \tag{7.111}$$

whose only non-vanishing commutation relation is

$$\left[\tilde{\mathbf{b}}_{\mathbf{q}}, \tilde{\mathbf{b}}_{\mathbf{q}}^{\dagger} \right] = 1. \tag{7.112}$$

The Hamiltonian becomes

$$\hat{H}_{plasma} = \hbar\omega_P \sum_{\mathbf{q}} \left(\tilde{\mathbf{b}}_{\mathbf{q}}^{\dagger} \tilde{\mathbf{b}}_{\mathbf{q}} + \frac{1}{2} \right). \tag{7.113}$$

Note that plasmons in the bulk oscillate with the plasma frequency determined by the free electron density and the effective electron mass.

7.10 Surface Plasmon Polaritons

Scientific and technological breakthroughs in the fabrication, observation, and characterization of the optical properties of nanostructures and metal-dielectric interfaces have opened new possibilities for manipulation and confinement of electromagnetic fields in very small spaces. In this section, we will discuss the quantization of electromagnetic evanescent waves propagating along a planar interface between a dielectric and a thin metallic film. As discussed in previous chapters, such coupled oscillations of the electron plasma and the EM field associated with it can be excited by electrons or by EM fields. Their energy quanta are the surface plasmons, sometimes known as surface plasmon polaritons (SPP) when reference is made both to the plasma oscillations within the metal and to the electromagnetic fields necessarily associated with them.

The frequency and wavevector of the SPP depend on the dimensions and properties of the metal and the dielectric. When the SPP is excited by an incident EM field, whose frequency and tangential component of the wavevector match those of the SPP, a surface plasmon resonance (SPR) occurs. SPRs can be used to obtain information on the properties of the interface and the dielectric.

In the next section, we will discuss the ability of surface plasmons to generate a local electromagnetic field around metal nano-objects or nanostructures embedded in the dielectric. The topology of an isolated nano-particle or that of a specific nano-structure array determines the characteristics of the SPP propagation modes, and hence of the possible surface plasmon resonances (SPR). A SPR localized in a nanostructure is called a localized surface plasmon resonance (LSPR).

Both resonances, SPR and LSPR, are extremely sensitive to subtle changes in the refractive index of the dielectric or in the geometry and size of the metallic nanostructures (isolated or arrays).

If losses in the metal are neglected, the quantization formalism of surface plasmon polaritons is analogous to that of C. K. Carniglia and L. Mandel (1971).[1] The electromagnetic field is quantized in modes that propagate in half semi-infinite dielectric and metal spaces and satisfy the classical Maxwell's equations with the appropriate boundary conditions at the interface. The metal is characterized by its dielectric function, which accounts for its dispersive properties originated via the collective response of the electrons.

7.10.1 Classical Surface Plasmon Polariton Waves at Planar Interfaces

Let us recall the classical solution for SPP waves as presented in Sect. 5.4. Since Maxwell's equations and boundary conditions preclude the existence of TE modes, we consider only TM modes.

The dielectric function takes different values in the two semi spaces:

$$\varepsilon(\mathrm{z}) = \begin{cases} \varepsilon_d & z > 0 \\ \varepsilon_m(\omega) & z < 0 \end{cases}. \tag{7.114}$$

For lossless waves propagating along $\hat{\mathbf{x}}$, the electric field given in (5.21b and c) and (5.23b and c) can be written as:

$$\mathbf{E}(x,z,t) = \begin{cases} E_0 \left[i\hat{\mathbf{x}} - \dfrac{\beta}{k_d}\hat{\mathbf{z}} \right] e^{i(\beta x - \omega t)} e^{-k_d z} + c.c., & z > 0 \\ E_0 \left[i\hat{\mathbf{x}} + \dfrac{\beta}{k_m}\hat{\mathbf{z}} \right] e^{i(\beta x - \omega t)} e^{k_m z} + c.c., & z < 0 \end{cases}. \tag{7.115}$$

We recall here that the frequency ω and the wavenumber β are linked by the dispersion relation (5.27). Notice that k_d and k_m characterize the exponential decay in the dielectric and metal respectively, as shown in Fig. 7.6. They are related to β by (5.22) and (5.24).

In general, SPP waves can propagate with any real wave vector $\boldsymbol{\beta}$ parallel to the interface and decay exponentially away from the interface in both half spaces, the dielectric and the metal.

Equation (7.115) can be written in compact form as

$$\mathbf{E}_{\boldsymbol{\beta}\pm}(\mathbf{r},t) = \frac{1}{\sqrt{\mathcal{L}}} \frac{k_\pm}{\omega \varepsilon_0 \varepsilon_\pm} \left[i\hat{\boldsymbol{\beta}} - \frac{\beta}{k_\pm}\hat{\mathbf{z}} \right] e^{i(\boldsymbol{\beta}\cdot\mathbf{r} - \omega_\beta t)} e^{-k_\pm z}. \tag{7.116}$$

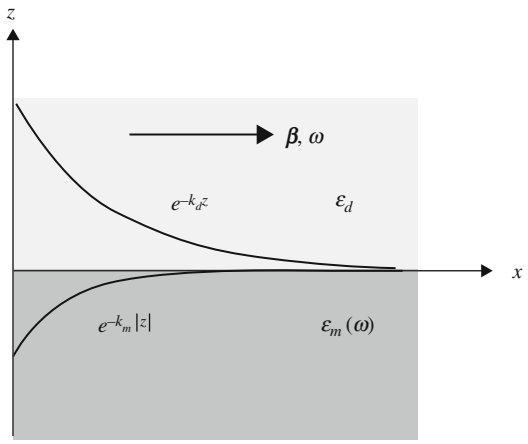


Fig. 7.6 Schematic representation of a surface plasmon wave, consisting of a wave that decays exponentially on both sides of the interface and propagates parallel to the surface with a real wave vector $\boldsymbol{\beta}$

with $\mathbf{E}_+(\mathbf{r},t)$, and $\mathbf{E}_-(\mathbf{r},t)$ being the SPP field for $z>0$ and $z<0$ respectively, α_0 a complex amplitude; $\varepsilon_+=\varepsilon_d$, $\varepsilon_-=\varepsilon_m$, $k_+=k_d$, and $k_-=-k_m$. $\mathcal{L}$ is a normalization length that will be obtained by normalization of the vector potential.

The SPP magnetic field for a given $\boldsymbol{\beta}$ is continuous at the interface and is given by

$$\mathbf{H}_{\boldsymbol{\beta}\pm}(\mathbf{r},\mathrm{t})=\frac{1}{\sqrt{\mathcal{L}}}\left(\hat{\mathbf{z}}\times\hat{\boldsymbol{\beta}}\right)e^{i\left(\boldsymbol{\beta}\cdot\mathbf{r}-\omega_\beta t\right)}e^{-k_\pm z}. \tag{7.117}$$

Both fields can be derived from the vector potential

$$\mathbf{A}_{\boldsymbol{\beta}\pm}(\mathbf{r},t)=\frac{1}{\sqrt{\mathcal{L}}}\frac{k_\pm}{\omega^2\varepsilon_0\varepsilon_\pm}\left[\hat{\boldsymbol{\beta}}+\frac{i\beta}{k_\pm}\hat{\mathbf{z}}\right]e^{i\left(\boldsymbol{\beta}\cdot\mathbf{r}-\omega_\beta t\right)}e^{-k_\pm z}. \tag{7.118}$$

Note that the fields $\mathbf{A}_{\boldsymbol{\beta}+}(\mathbf{r},t)$ and $\mathbf{A}_{\boldsymbol{\beta}-}(\mathbf{r},t)$ are zero for $z\le 0$ and $z\ge 0$, respectively.

The full description of the SPP requires both fields $\mathbf{A}_{\boldsymbol{\beta}\pm}(\mathbf{r},t)$, which are not independent. They are the solution of the boundary value problem of Maxwell's equations within an infinite volume with half-space dielectric and half-space metal and can be considered as a duplet mode:

$$\mathbf{A}_{\boldsymbol{\beta}}(\mathbf{r},t)=\mathbf{A}_{\boldsymbol{\beta}+}(\mathbf{r},t)\vartheta(z)+\mathbf{A}_{\boldsymbol{\beta}-}(\mathbf{r},t)\vartheta(-z) \tag{7.119}$$

where $\vartheta(z)$ is the Heaviside theta step function

$$\vartheta(z)=\begin{cases}1 & z>0\\ 0 & z<0\end{cases}. \tag{7.120}$$

The vectors $\mathbf{A}_{\boldsymbol{\beta}}(\mathbf{r},t)$ satisfy the orthonormality condition (7.74) with the normalization length

$$\mathcal{L}=\frac{8\pi^2\beta^2}{\hbar\omega_\beta^3\varepsilon_0}\left(\frac{1}{k_+\varepsilon_+}+\frac{1}{k_-\varepsilon_-}\right). \tag{7.121}$$

7.10.2 *Quantization of the SPP Waves at Planar Interfaces*

For the quantization, we follow the procedure described in Sect. 7.7. We expand the operator $\tilde{\mathbf{A}}(\mathbf{r},t)$ in terms of the orthonormal modes as

$$\tilde{\mathbf{A}}(\mathbf{r},t)=\sum_{\boldsymbol{\beta}}\left[\mathbf{A}_{\boldsymbol{\beta}}(\mathbf{r},t)\tilde{a}_{\boldsymbol{\beta}}+\mathbf{A}^*_{\boldsymbol{\beta}}(\mathbf{r},t)\tilde{a}^\dagger_{\boldsymbol{\beta}}\right] \tag{7.122}$$

where $\hat{a}_{\boldsymbol{\beta}}$ and $\hat{a}^{\dagger}_{\boldsymbol{\beta}}$ are the annihilation and creation operators for photons in the SPP duplet mode $\boldsymbol{\beta}$ given in (7.119). They satisfy the bosonic commutation relations

$$\left[\tilde{a}_{\boldsymbol{\beta}}, \tilde{a}^{\dagger}_{\boldsymbol{\beta}'}\right] = \delta_{\boldsymbol{\beta}\boldsymbol{\beta}'} \tag{7.123a}$$

$$\left[\tilde{a}^{\dagger}_{\boldsymbol{\beta}'}, \tilde{a}^{\dagger}_{\boldsymbol{\beta}}\right] = \left[\tilde{a}_{\boldsymbol{\beta}'}, \tilde{a}_{\boldsymbol{\beta}}\right] = 0. \tag{7.123b}$$

There is no sum over polarizations in (7.122) since SPP waves exist only as TM modes, and their polarization is well defined for each $\boldsymbol{\beta}$.

Since the normal modes $\mathbf{A}_{\boldsymbol{\beta}}(\mathbf{r}, t)$ are monochromatic and satisfy the orthonormality condition (7.74), the Hamiltonian of the SPP waves can be written as the sum of independent harmonic oscillator Hamiltonians:

$$\tilde{H} = \sum_{\boldsymbol{\beta}} \hbar\omega_{\boldsymbol{\beta}} \left(\tilde{a}_{\boldsymbol{\beta}} \tilde{a}^{\dagger}_{\boldsymbol{\beta}} + \frac{1}{2}\right). \tag{7.124}$$

The total energy corresponds to the sum of the energies of independent harmonic oscillators, each associated with a mode $\boldsymbol{\beta}$.

7.10.3 Modal Functions and Source Current Densities

In the previous section we worked with the transverse fields which are associated only with transverse currents in the metal. Y. O. Nakamura[30] performed the quantization of SP waves using the hydrodynamic jellium model for electron gas in the metal. Maxwell's equations (1.a) and (1.d) in the metal establish the link between SP waves and the current density waves in the metal associated with them:

$$\nabla \cdot \mathbf{E} = eN \tag{7.125a}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{E}}{\partial t} + eN_0 \boldsymbol{v}, \tag{7.125b}$$

where N and $\boldsymbol{v}$ are fluctuations of the electron density and average velocity of electrons, e is the electron charge, and N_0 is the electron density in the undisturbed state of the electron gas.

In the hydrodynamic jellium model of collective electron oscillations, the electron current density fluctuations obey the following continuity equation:

$$N_0 m \frac{\partial \boldsymbol{v}}{\partial t} = eN_0 \mathbf{E} - \frac{3mv_F^2}{5} \nabla N, \tag{7.126}$$

[30] Nakamura YO (1983) Prog Theor Phys. 70:908

where m is the electron mass and v_F is the Fermi velocity. Introducing the scalar and vector potentials, ϕ and $\boldsymbol{A}$, in the Coulomb gauge, we can decompose the electric field into transverse and longitudinal components that obey

$$\mathbf{E}_T = -\frac{\partial \boldsymbol{A}}{\partial t} \tag{7.127a}$$

$$\mathbf{E}_L = -\nabla\phi. \tag{7.127b}$$

The transverse and longitudinal components of the vector potential and velocity obey

$$\frac{\partial^2 \boldsymbol{A}}{\partial t^2} - \nabla^2 \boldsymbol{A} = eN_0 \boldsymbol{v_T} \tag{7.128a}$$

$$m\frac{\partial \boldsymbol{v_T}}{\partial t} = -e\frac{\partial \boldsymbol{A}}{\partial t} \tag{7.128b}$$

$$\nabla^2\phi = -eN \tag{7.128c}$$

$$\nabla\left(\frac{\partial\phi}{\partial t}\right) = eN\boldsymbol{v}_L \tag{7.128d}$$

$$N_0 m\frac{\partial \boldsymbol{v}_L}{\partial t} = -eN_0\nabla\phi - \frac{3mv_F^2}{5}\nabla N \tag{7.128e}$$

The solutions of (7.128a) and (7.128b) lead to the expression (7.119) for the vector potential, with $\boldsymbol{v_T} = -e\boldsymbol{A}/m$ in the metal and zero elsewhere. The modal expansions for the scalar potential, the electron density waves, and the longitudinal velocity in the metal $(z<0)$ are given by

$$\phi(\mathbf{r},t) = \sum_{\boldsymbol{\beta}} \phi_{\boldsymbol{\beta}} e^{i\boldsymbol{\beta}\cdot\mathbf{r}+\gamma z} \tag{7.129a}$$

$$n(\mathbf{r},t) = \sum_{\boldsymbol{\beta}} \frac{\gamma^2-\beta^2}{e}\phi_{\boldsymbol{\beta}} e^{i\boldsymbol{\beta}\cdot\mathbf{r}+\gamma z} \tag{7.129b}$$

$$\boldsymbol{v}_L(\mathbf{r},t) = \hat{\boldsymbol{\beta}}\sum_{\boldsymbol{\beta}} \frac{(i\beta+\gamma)}{e}\dot{\phi}_{\boldsymbol{\beta}} e^{i\boldsymbol{\beta}\cdot\mathbf{r}+\gamma z} \tag{7.129c}$$

$$\gamma^2 = \beta^2 + \frac{5}{3}\frac{\omega_P^2-\omega^2}{v_F^2}. \tag{7.129d}$$

By requiring specular reflection at the interface, i.e., that the normal component of the electronic velocity vanish at $z=0$, and after applying the boundary

conditions for the electric and magnetic fields, all matter fields in the metal and evanescent waves in the dielectric can be expressed in terms of the ϕ_{β} canonical coordinates and their conjugate momenta. Canonical quantization and appropriate normalization lead to the quantized energy of (7.124). All fields, including the scalar potential, the electron density and velocity fluctuations can be expressed in terms of the same creation and annihilation operators for SPP, $\tilde{a}^{\dagger}_{\beta}$ and $\tilde{a}_{\beta}$ defined in (7.123a and b). This quantization procedure highlights the quantum nature of the combined electromagnetic field and electron density oscillations.

Both quantization procedures, the former and that of Sect. 7.10.2, rely on modal expansions for stationary monochromatic fields. A. Archambault et al.[31] derived two equivalent vectorial representations of a surface plasmon field for arbitrary source currents, using an expansion over surface waves with either a complex wave vector or a complex frequency. The two representations lead to different shapes in the dispersion relation, as shown in Fig 7.7. The representation using complex wave vectors is well adapted to stationary monochromatic fields, in a source-free region, i.e., beyond the source region. It leads to a dispersion relation, shown in Fig. 7.7a, that presents a back- bending at a cut-off value of the real part of the SPP wave vector. The representation using complex frequencies is well adapted to SPP fields excited by pulses after the excitation has died off. Its corresponding dispersion relation, shown in Fig 7.7b, is cut-off free.

The back-bending towards the light line near the surface plasmon energy was first observed by Arakawa et al.[32] upon measuring the dispersion by means of attenuated total reflection, varying the angle of incidence at fixed frequency. The back-bending means that the SPP does not exist beyond a cut-off momentum. They observed, however, that the dispersion relation obtained from a spectrum at fixed angle did not backbend.

The existence of a cutoff momentum is relevant to the applications of SPP relying on strong field confinement like optical lithography, enhanced nonlinear effects and super-resolution imaging. A cutoff spatial frequency implies an upper limit to resolution or confinement achievable with SPP.

The general form of the surface plasmon field, in the absence of loss, was derived by A. Archambault et al. (2009)[30] from the Green's tensor for an arbitrary source current density $\boldsymbol{j}(\mathbf{r}, t)$ located near the dielectric-metal interface. For purposes of the quantization of the SPP field, the more suitable representation for the problem at hand can be chosen. The representation with real frequencies and complex wave vectors provides normal modes appropriate for the quantization of the electromagnetic field consistent with the expansion used in the preceding section. Note also that in the presence of losses, the sources must be active to maintain a stationary regime.

[31] Archambault A et al (2009) Phys Rev. B79:195414

[32] Arakawa ET et al (1973) Phys Rev Lett. 31:1127

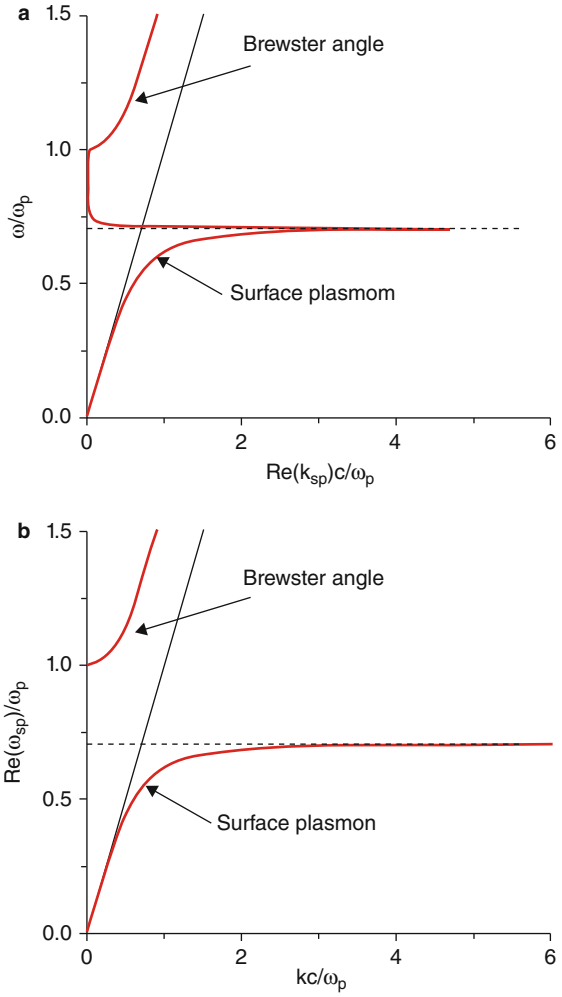


Fig. 7.7 Drude's dispersion relation of surface plasmons propagating along the metal/air interface. (**a**) Representation using real frequencies. (**b**) Representation using real wave vectors. (From Archambault et al. 2009)

7.10.4 The Quantization of the SPP Fields in Other Geometries

Metal–dielectric interfaces of arbitrary geometries also support charge density oscillations like the surface plasmons characteristic of planar interfaces. SPP are bound to the metal surface in the perpendicular direction and propagate non-radiatively along the surface. This behavior is inherent to the boundary conditions at the interface and persists for interface geometries other than planar. Consequently, SPP can be guided through multiple surfaces, bends,[33] rectangular strips, v-grooves, waveguide

[33]Rossouw D, Botton GA (2013) Phys Rev Lett. 110:066801

arrays, nanowires, etc.[34] The quantization procedure presented in Sect. 7.7 can be carried out for these more complex geometries, by determining the mode functions $\mathbf{A_K}(\mathbf{r})$ that solve the Maxwell equations in those geometries, along with the corresponding boundary conditions. The quantization is usually performed in a finite volume, i.e., in terms of a discrete set of modes, although in some applications the continuum limit for the vector $\mathbf{K}$ is used.

7.10.5 Quantization of SPP Fields in Lossy Media

The canonical quantization of the electromagnetic field in dispersive and lossy dielectrics was performed by B. Huttner and S. M. Barnett[35] with a microscopic model in which the quantum system is represented by the free electromagnetic field, the polarization field in the medium, and a reservoir that accounts for losses. The medium is modelled as a harmonic oscillator with a resonance frequency ω_0, and the reservoir consists of an ensemble of harmonic oscillators with a continuum frequency spectrum. The polarization of the medium interacts with both the free EM field and the bath of harmonic oscillators. This rigorous method results in space-dependent creation and annihilation operators that satisfy a set of equal-space commutation relations and obey spatial Langevin equations of evolution. The coupling to a reservoir of bath modes is currently used to model losses for SPP waves. The strength of the coupling to the reservoir is derived phenomenologically from the imaginary part of the dielectric function $\varepsilon(\omega)$ for the metal that arises from the damping experienced by the electrons in the metal.

7.11 Localized Surface Plasmon Resonances

Nanoparticles of noble metals that can support plasmon oscillations are known as plasmonic nanoparticles. The excitation of the free electrons in the nanoparticle by an EM field is characterized by a local excitation in the surface of the nanoparticle rather than by the propagation of a wave along a metal-dielectric interface. Hence, the geometry of the nano-object establishes the key features of its interaction with EM fields. The quanta of the corresponding plasmon modes are known as localized surface plasmons (LSP). When an incident field excites resonantly a LSP mode of the nanoparticle, we have a localized surface plasmon resonance (LSPR), which provides extraordinary optical properties to the host material. The confinement of plasmons in nano-objects, whose dimensions are less than the wavelength of the EM field, leads to enhancement of the local electric field about the nano-object. The

[34]Barnes WL et al (2003) Nature 424:824; Maier SA, Atwater HA (2005) J Appl Phys. 98:011101; Ozbay E (2006) Science 311:189

[35]Huttner B, Barnett SM (1992) Phys Rev. A46:4306

enhancement decays rapidly with distance away from the nano-object. LSPRs can be tailored by changing the size and shape of the nanoparticles and can be used to enhance luminescence of a quantum emitter embedded in the dielectric, to enhance nonlinear processes, to confine light in nanoscale regions, etc. These effects have great potential for applications in optical communications, as well as for nanoantennas for optical probes in chemistry, biology, and medicine.

Coming back to our quantization procedure, LSP modes are quite different from bulk plasmon and surface plasmon modes. Bulk plasmonic oscillations are longitudinal electron density waves in the whole metal volume, surface plasmons are electron density waves traveling along an interface, and localized surface plasmons are confined to the surface of a nanoparticle. Both SP and LSP involve coupling of the electric charge density on the metal's surface with the electromagnetic field that excites the plasmon oscillation.

Quantization of the LSP is usually performed by determining the normal oscillation modes of the plasmonic nanostructure that solve Maxwell's equations for the interaction of EM waves with the nanostructure. Assuming for example that the EM field is a plane wave and the nanostructure is a homogeneous metallic nanosphere, we would obtain analytical solutions for the field inside and outside the nanosphere by means of Mie Scattering theory. The solution of Maxwell's equations for other geometrical shapes may require however the use of numerical methods.

When the frequency of incident light matches the natural frequency of a LSP mode, a localized surface plasmon resonance occurs. The spatial structure and natural frequency of the LSP mode are extremely sensitive to subtle changes in the dielectric and to the size and geometry of the metallic nanostructure. This sensitivity allows for fine tailoring of the LSPR and are the basis of many of the applications described in the preceding chapter.

Localized surface plasmons often arise from a localized scatterer inside a homogeneous medium illuminated by incoming light. A central problem for a quantum mechanical treatment is the appropriate normalization of the electromagnetic field. The usual recipe for quantization for a specific problem, as given by W. E. Lamb,[36] starts by deciding "how much of the universe needs to be brought into the discussion," which involves the consideration of some finite or periodic quantization volume with specific boundary conditions for the classical problem, and proceed to the determination of the normal modes of the classical field in that volume and their adequate normalization in the absence of a source. Thereafter one decides how to model the light sources and work out how they drive the system.

The eigenmodes of nano or plasmonic structures are solutions to Maxwell's equations satisfying all boundary conditions for the given material distribution and geometry. The quantum treatment of light-matter interactions requires the appropriate normalization of electromagnetic fields. As remarked recently by J. Oppermann et al.,[37] the normalization process as described in Sect. 7.7 for

[36]Lamb WE, Jr. (1995) Appl Phys. B66:77

[37]Oppermann J et al (2017) arXiv:1702.00214

plasmonic structures and quantum cavities, is usually not obvious since boundary conditions at the interfaces can lead to δ-distributions within the infinite spatial extent of the normalization integral (7.79).

Following the trend of the pioneering work of Carniglia and Mandel (1971),[1] J. Oppermann et al. provided a novel normalization approach for the scattering modes that arise when a localized scatterer of finite size is illuminated by a field. The scatterer can be encased in a virtual sphere of radius R. Maxwell's equations are solved in the homogenous medium outside the scatterer and the derived normalization conditions depend only on the electromagnetic field on the surface of a sphere containing the scatterer. The key achievement of their approach is to replace the normalization integration across the infinite space by surface integrations, an approach that can be used for localized metallic and dielectric scatterers of different geometries. This approach allows normalization to be treated in a numerically feasible and rigorous way, and is especially useful when considering strong field enhancement in localized plasmonic systems.

7.12 Absorption of Evanescent Photons and Stimulated Emission of Surface Plasmons

The interaction at the microscopic level of evanescent waves with atomic dipoles leading to the absorption and emission of evanescent photons was first discussed by Carniglia, Mandel, and Drexhage (1972),[28] who studied the excitation probability of dye molecules by evanescent light. They also described the inverse process, the spontaneous emission of evanescent waves by excited atoms or molecules. For this purpose, they considered classical fields since spontaneous emission takes place in the continuum of free-space electromagnetic modes. Their experiments and theoretical results obtained using the triplet modes lead to the conclusion that evanescent photons can be emitted and absorbed like homogeneous photons.

The process of light amplification by indirect coupling to an active medium through an evanescent wave, a method commonly used to pump Erbium doped fiber amplifiers in optical telecommunication systems, was studied by Chang-Woo Lee et al.[38] They developed a quantum theory of stimulated emission from an atom coupled to an evanescent field using the quantized triplet modes of the electromagnetic field and calculated the gain coefficients as a function of the atom distance to the interface for several incident angles. Large variations of the gain close to the interface were observed, as well as a strong decrease of the gain at the critical angle, due to an increase of the spontaneous emission rate. The inverse process, the emission of evanescent waves by excited atoms or molecules, was also described in Carniglia et al. (1972), but by means of classical theory, since the spontaneous emission takes place into the continuum of free-space electromagnetic modes.

[38]Chang-Woo L et al (2000) Phys Rev. A62:053805

The key property of engineered metallic nanostructures is their ability to concentrate light. However, there is drawback: metal-light interaction is always accompanied by non-negligible resistive heating losses that limit strongly the possibility of using plasmonic waveguides. The stimulated emission of SPP has hence awakened considerable interest, since gain could compensate absorption and radiation losses. Stimulated emission of SPP at the interface between a silver film and an optically pumped dye solution acting as the active medium was demonstrated with an attenuated-total-reflection configuration.[39] A. Archambault et al.[40] developed the quantum theory of spontaneous and stimulated emission of surface plasmons propagating along a plane interface, using the quantized SPP modes described in Sect. 7.10.2.

Spasers, as discussed in Chap. 6, are another application of the coupling between a set of quantum emitters which transfer their excitation energy to a nanostructure able to support localized plasmons of the same frequency, in a non-radiative resonant energy transfer process.

A simple theoretical model of the spaser, proposed by E. S. Andrianov et al.,[41] consists of a two-level system placed near a plasmonic spherical nanoparticle. The energy from the pumped two-level system is non-radiatively transferred to the nanoparticle, exciting surface plasmons. Spaser action is achieved when the Joule losses in the nanoparticle are compensated by the pumping. The quantum theory of a spaser-based nanolaser was developed by V. M. Parfenyev and S. S. Vergeles.[42] Their model considers a large number of identical two-level systems resonantly coupled to a low-Q plasmon cavity.

7.13 Plasmons and Quantum Information

Photonics is the foundation of fiber optics communications, as electronics is of computing. Plasmonics opens the possibility of interfacing photonics and nanoelectronics, thereby enabling the routing and manipulation of light at the nanoscale.[43] Quantum plasmonics and future integrated quantum technologies based on SP have great potential for miniaturization and scalability, but applications in quantum computing and quantum communications will require efficient photon-SPP-photon conversion and preservation of the quantum properties of light during the complete process. We discuss the advances in these two areas in the following subsections.[44]

[39]Seidel J et al (2005) Phys Rev Lett. 94:177401

[40]Archambault A et al (2010) Phys Rev. B82:035411

[41]Adrianov ES et al (2013) Opt Express 21:13467

[42]Vergeles SS, Parfenyev VM (2014) Opt Express 22:13671

[43]Shalaev VM, Brongersma ML (2010) Science 328:440

[44]A general survey of single photon sources and related problems may be found for example in Bertolotti M et al (2015) Progr Opt. 60:1

7.13.1 The Coupling of Photons to SPP

At the beginning of the 21st century, W.L. Barnes et al.[45] were concerned with the problem of efficient collection of the photons produced by solid-state single photon sources, and proposed to couple the emitter to a surface plasmon mode and then couple the surface plasmon mode back to a radiation mode. A few years later A. V. Akimov et al.[46] demonstrated direct coupling of the emission of a CdSe quantum dot to guided surface plasmons in a silver nanowire. When the quantum dot is optically excited in close proximity to the metallic nanowire, the wire's ends light up. The surface plasmons seem to capture the majority of the spontaneous radiation of the quantum dot into the guided surface plasmon modes over a broad band of optical frequencies, since the strong coupling occurs far away from the plasmon resonance frequency and does not involve an optical cavity. The quantum dot emits only one photon at a time, and photon-coincidence measurements between the free-space fluorescence of the quantum dot and scattering from the nanowire's end showed that the quantum dot can either radiate the single photon into free space or into the surface plasmon modes, but never both simultaneously.

Nitrogen vacancies in diamond crystal lattices are a novel light source that operate in a very broad spectrum and at room temperature. By using these color centers in single diamond nanocrystals as a single-photon source, R. Kolesov et al.[47] demonstrated that surface plasmon polaritons in a silver nanowire exhibit both wave and particle properties. The measured second-order correlation (Hanbury Brown–Twiss experiment) between the fluorescence at the two ends of the wire showed no counting coincidences at zero delay time. This observation of anti-bunching provided the experimental evidence that the plasmons originate from the fluorescence of a single quantum system, and then behave as single quanta of the LSP wave. This observation is remarkable since LSP are associated with electron charge density waves propagating along a dielectric-metal interface. The wavelike behavior of the LSP was also demonstrated by recording the self-interference of the LSP. Since the quantum emitter is likely to excite SP in either direction, the plasmon propagating in one direction will interfere with the plasmon emitted in the opposite directions after reflection off the opposite end of the nanowire. The interference patterns observed at both ends of the nanowire are consistent with those expected from the light output of a lossy cavity.

A clear manifestation of SPP wave particle duality and a visualization of plasmonic fields at the nanoscale was provided in Piazza et al.[2] with the simultaneous observation of the spatial interference and the quantization of confined plasmonic fields by means of ultrafast transmission electron microscopy. Electron and light pulses were overlapped on a single nanowire suspended in a graphene film. SPPs in the metallic nanowire were photo-excited and probed by relativistic electrons

[45]Barnes WL et al (2002) Eur Phys J. D18:197

[46]Akimov AV et al (2007) Nature 450:402

[47]Kolesov R et al (2009) Nature Phys. 5:470

photoemitted from a cathode by femtosecond laser pulses. The delay between the two pulse trains was controlled via an optical delay line, allowing for time-resolved optical-pump/electron probe experiments. The energy exchange between single electrons and the quanta of the photoinduced near-field was imaged synchronously with the spatial interference pattern of the LSP field in the nanowire. A map of the electron energy loss intensity versus the relative time delay Δt between the optical pump and electron probe pulses is shown in the upper part of Fig. 7.8. The energy spectrum for $\Delta t = 0$ shows equally spaced quantum jumps that reflect the quantized energy of the SPP. A characteristic Fabry-Perot interference pattern consistent with the one-dimensional confinement of the field in a nanowire is in the lower part of Fig. 7.8, showing simultaneously the wave properties of SPP.

This methodology enables not only the visualization but also the control of plasmonic fields at the nanoscale, and constitutes a promising tool for the development of advanced photonic circuits, and the enhanced understanding of the fundamental properties of confined electromagnetic fields.

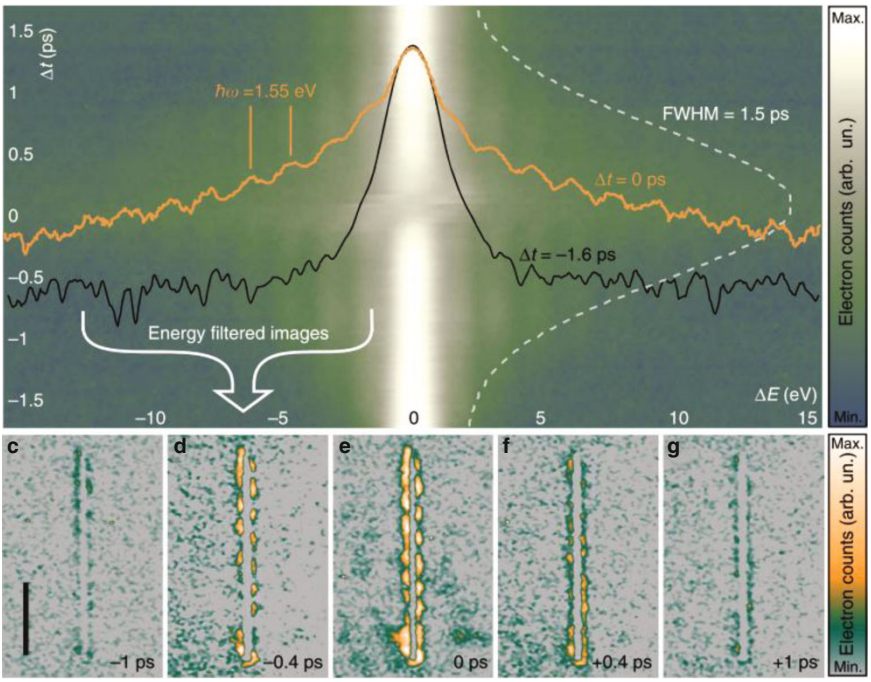


Fig. 7.8 Map of the electron energy loss intensity versus the relative time delay Δt between the optical pump and electron probe pulses in log scale. Energy spectra, also in log scale, at $\Delta t = -1.6$ ps and $\Delta t = 0$ are superimposed in black and orange traces respectively. Inserts (c-g) show snapshots of an isolated nanowire at different time delays obtained using only the electrons that have gained energy, i.e., those in the region indicated by the white arrow. Electron counts are on a linear scale. The vertical scale bar in c corresponds to 2 μm and holds for all images.
(From Piazza L et al. 2015)

7.13.2 Preservation of the Properties of a Quantum State Under Conversion to SPP and Back

Quantum information processing requires not only the possibility of coupling quantum emitters to plasmonic modes but that the excited SPPs preserve the quantum statistical properties of the source and their non-classical character under propagation. M. S. Tame et al.[48] studied the excitation of SPP by single photons, using an attenuated reflection set-up. The optimal photon-SPP coupling is a function of the frequency and was found to be remarkably efficient. They also studied to what extent the excited SPPs preserve the quantum statistics of the photons as they travel on realistic metal surfaces, and found that the entire process of photon-to-SPP transfer and SPP propagation is analogous to single photon propagation through an array of lossy beam splitters.

A. Huck et al.[49] demonstrated the efficient generation, propagation, and reemission of squeezed long-range SPP in a gold waveguide. The squeezed SPP was excited by squeezed light, and both the incident and reemitted light were characterized by means of quantum tomography to determine their degree of squeezing. Despite linear loss and decoherence in the plasmonic mode, the squeezing was coherently transferred from the incident light to the SPP and back to the reemitted light, the quantum state of the last being degraded solely by vacuum noise.

The building blocks of quantum information are the qubits. Today's computers work with binary digits called bits, which only take "0" or "1" values. Unlike the classical bit, a qubit has a probability of being either in "0" or "1" at the same time, which is the key element that differentiates classical logics from the logics of quantum computing. Furthermore, the quantum state of multiple qubits can also be entangled, i.e., the quantum state cannot be written as the direct product of states of individual qubits. Entanglement is one of the most fundamental phenomena of quantum theory, is the key element for quantum computing, and has no classical counterpart. A simple example of entangled quantum state is the quantum state of polarization of bi-photons produced by parametric down conversion in a nonlinear crystal, a process whereby a single photon splits spontaneously into two polarization entangled photons described by the quantum state:

$$|\Psi\rangle = \frac{1}{2}(|H_1, V_2\rangle + |V_1, H_2\rangle). \tag{7.130}$$

The bi-photon state $|\Psi\rangle$ is a superposition of two bi-photon states that have the same probability: photon 1 with horizontal polarization and photon 2 with vertical polarization ($|H_1, V_2\rangle$), or photon 1 with vertical polarization and photon 2 with horizontal polarization ($|V_1, H_2\rangle$). Quantum mechanics does not predict which of

[48] Tame MS et al (2008) Phys Rev Lett. 101:190504

[49] Huck A et al (2009) Phys Rev Lett. 102:246802

the two states is going to be the outcome of a measurement; it only predicts that each state has the same probability to be the outcome of the measurement.

Entanglement of a large number of qubits, and its preservation for a time long enough to perform quantum operations and record their outcome, are the key issues for quantum computation. Survival of entanglement over long-haul transmission of information is also at the core of quantum cryptography. Preservation of entanglement after photon-plasmon-photon conversion has been observed in two different experiments. E. Altewischer, M. P. van Exter and J. P. Woerdman[50] investigated the effects of inserting a plasmonic device in each of the paths of polarization-entangled bi-photons. A nanostructured metal film obtained by perforating the film with a periodic array of subwavelength holes was used to convert photons into SP waves, which after tunneling through the holes, reradiate free-space propagating photons. Coincidence count measurements showed that the quantum entangled state maintained almost completely its high purity through the transition from photon to SP and back. The preservation of the entanglement in plasmon-assisted light transmission was corroborated by S. Fasel et al.,[51] who transmitted energy-time entangled twin photons[52] through a similar subwavelength metallic hole array, and through a half centimeter long gold stripe waveguide.

Current nanofabrication technologies allow the engineering of the dielectric constant with metamaterials, and the tailoring of the dispersion relation of homogenous media into specific optical properties like those of hyperbolic materials,[53] where light perceives the space as metallic in one direction and dielectric in another. The possibilities for increased interactions with quantum emitters and channeling of light in these new materials are under exploration.

[50] Altewischer E et al (2002) Nature 418:304

[51] Fasel S et al (2005) Phys Rev Lett. 94:110501

[52] Brendel J. et al (1999) Phys Rev Lett. 82:2594

[53] Jacob Z et al (2011) Science 334:463

Index

A
Amplitude reflection coefficient, 41, 44
Anomalous dispersion, 7
ATR, 141
  sensors, 108–109
Attenuated total reflection, 54

B
Backward wave, 3
Band gap, 105
Beam waist, 31
Beer's law, 7
Biaxial crystals, 28–29
Black-body, 202–208
Bloch phase, 99
Brewster angle, 42
Brillouin zones, 100
Buried waveguide, 80
Butt coupling, 81

C
Cartesio-Snell law, 35, 55
Channel Plasmon polaritons CPP), 184–186
Charge density, 1
Chirality, 29–30
  parameter, 29
Cladding, 69, 88
Classical resolution limit, 62–64
Coherent states, 217–219
Conductivity
  dynamic, 12, 13
  electric, 2
CPP, 184–185
Current density, 1
Cut-off, 76

D
Dichroic mirrors, 29
Dichroism, 29
Dielectric constant
  relative, 6
Dielectric displacement, 1
Dielectric tensor, 23
Dispersion, 5–9
  anomalous, 7
  normal, 7
  number of dispersion electrons, 8
Dispersive media, 5
Drude model, 9
Dyakonov waves, 163–166

E
Electric
  conductivity, 2, 12
  field, 1
  permittivity, 2
  relative permittivity, 2
  susceptibility, 2
Enantiometers, 29
End-fire coupling, 81
Epi (subcritical angle illumination)
  fluorescence microscopy, 123
Evanescent field
  definition, 47
Evanescent waves, 35, 45, 112
  in diffraction, 65–66
  energy transport by, 49–50
  in optical waveguides, 69
  X-rays, 57–59
Exciton polaron, 156
Extraordinary index, 26

© Springer International Publishing AG 2017

M. Bertolotti et al., *Evanescent Waves in Optics*, Springer Series in Optical Sciences,
DOI 10.1007/978-3-319-61261-4

F
Far field, 112
Fast light, 15
FEWS, 109
Fibres
  core, 88
  graded index, 89
  holey, 109
  microstructured, 109
  step index, 89
Fock states, 215–217
Forward wave, 3
Fresnel coefficients, 39–45
  formula for negative index materials, 57

G
Gaussian beams, 31
  beam waist, 31
  Rayleigh length, 32
  reflection and refraction of, 64–65
Goos-Haenchen shift, 48, 50
Gouy phase shift, 33
Group velocity, 5, 14–16

H
Helmholtz equation, 3

I
Impedance
  of medium, 4
  of vacuum, 5
Incidence angle, 36
  plane, 36
Index ellipsoid, 24
Inhomogeneous field, 47
Intensity reflection coefficient, 41

J
Jones matrices, 19–23

L
Left-handed materials, 56
Localized surface plasmon, 150
Longranging plasmon polaritons (LRSPP), 147

M
Magnetic
  field, 1
  induction, 1
  permeability, 2
Magnetization, 1
Maxwell equations, 1
Metamaterials, 57
Microscope
  fluorescence (TIRF), 123
  Scanning Near-field Optical (SNOM), 112
  Scanning Tunnelling Optical (STOM), 118
m-lines, 85

N
Nanolaser, 192
Near-field microscopes, 48, 50, 111
Negative index materials, 55, 166
Normal dispersion, 7
Number of dispersion electrons, 8

O
Objective lens method in TIRF, 123
Optical activity, 29
Optical axis, 26
Optically anisotropic media, 23–29
Ordinary index, 26

P
Passing bands, 99
PBG, 96, 100–106
Phase-matching condition, 83
Phase velocity, 4
Phonon polaritons, 158
Photonic band gap, 69, 96
  crystals, 96
Photons, 209, 213, 214, 227
Plane wave decomposition, 60–62
Plasma frequency, 10
  oscillations, 127
Plasmon microscope, 201–202
Plasmon polaritons, 128
Plasmons, 127, 209, 239
  bulk, 130–131
  dispersion relation, 136–139
  in graphene, 168
  localized surface, 150–155
  surface Plasmon polaritons, 127, 131–136
Poincarè sphere, 22
Polaritons, 157
Polarizability, 6
Polarization of light
  anticlockwise polarized light, 16, 18
  circularly polarized light, 16
  clockwise polarized light, 16, 18
  direction of, 16

elliptically polarized light, 16
linearly polarized, 16, 18
plane of, 16
Polarization vector, 1
nonlinear, 2
Poynting vector, 4
Principal axis, 23, 26
Prism method of coupling, 123
Propagation constant, 3

Q
Quanta, 209
Quantization of evanescent waves, 209–255

R
Radiation modes, 79
Raman effect, 170
surface enhanced (SERS), 170–172
Rayleigh length, 32
Reflection angle, 36
law, 36
Refraction angle, 36
law, 36
Refractive index, 2
Relative permeability, 2
Relative permittivity, 2
Relaxation time of electrons, 10
Resolution, 67, 174
Resonant condition, 171
Ridge waveguide, 80

S
Sensors
with evanescent waves, 106
plasmonic, 106, 173–176
SERS, 170–172
Short ranging Plasmon polaritons, 147
Slow light, 15
SNOM, 112–118
Solar cells, 198–201
Spaser, 193, 211
SRSPP, 147
STED, 111
Stokes parameters, 19–23
STOM, 111, 118–122
Stopping bands, 99
Substrate radiation modes, 79
Surface plasmon, 137
frequency, 137
polariton, 127
resonance sensors, 173

T
Tamm plasmon polaritons, 167
TIR, 47
TIRF, 123–126
Topological insulator, 13
Total internal reflection, 47, 49
fluorescence, 123
reflection for X-rays, 58
Transmission coefficient, 177
Tunnelling, 50

U
Uniaxial media, 26–27

W
Waveguide, optical, 69
planar, 69–88
Wave number, 3
vector, 3
Weak coupled theory, 86